REVIEW

DNA information: from digital code to analogue structure

BY A. A. TRAVERS1,2,*, G. MUSKHELISHVILI3 AND J. M. T. THOMPSON4,5

1 MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK
2 Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK
3 School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany
4 School of Engineering, University of Aberdeen, Aberdeen, UK
5 Department of Applied Mathematics and Theoretical Physics, Cambridge, UK

The digital linear coding carried by the base pairs in the DNA double helix is now known to have an important component that acts by altering, along its length, the natural shape and stiffness of the molecule. In this way, one region of DNA is structurally distinguished from another, constituting an additional form of encoded information manifest in three-dimensional space. These shape and stiffness variations help in guiding and facilitating the DNA during its three-dimensional spatial interactions. Such interactions with itself allow communication between genes and enhanced wrapping and histone–octamer binding within the nucleosome core particle. Meanwhile, interactions with proteins can have a reduced entropic binding penalty owing to advantageous sequence-dependent bending anisotropy. Sequence periodicity within the DNA, giving a corresponding structural periodicity of shape and stiffness, also influences the supercoiling of the molecule, which, in turn, plays an important facilitating role. In effect, the super-helical density acts as an analogue regulatory mode in contrast to the more commonly acknowledged purely digital mode. Many of these ideas are still poorly understood, and represent a fundamental and outstanding biological question. This review gives an overview of very recent developments, and hopefully identifies promising future lines of enquiry.

Keywords: DNA information; DNA stiffness; DNA bending; DNA supercoiling; chromosome structure

1. Introduction

The three-dimensional organization of DNA genomes, produced by global deformations of the DNA chain, plays an important, yet poorly understood,

*Author for correspondence (aat@mrc-lmb.cam.ac.uk).

One contribution of 14 to a Theme Issue ‘Beyond crystals: the dialectic of materials and information’.
role in gene regulation. In the nucleus of eukaryotic cells, as well as in the nucleoid of bacteria, DNA is organized in a complex superstructure that facilitates communication between genes that are either on different chromosomes or are widely separated from each other on a linear map of the same chromosome. Understanding how this organization is established is a fundamental, outstanding, biological question. We argue that the digital base sequence encodes information specifying not only protein sequences but also the three-dimensional trajectory of the DNA chain itself.

Although it is known that, in the nucleus, individual chromosomes occupy distinct non-overlapping domains, or chromosomal territories [1], the molecular details of large protein–DNA complexes in eukaryotic chromatin are essentially unknown. An outstanding exception is the structure of the nucleosome core particle (NCP), the fundamental unit of chromatin structure, as revealed by several crystal structures whose resolution now extends to 1.9 Å [2–6]. In a typical NCP, approximately 145 bp of DNA are wrapped in a left-handed superhelical spiral around an octameric assembly containing two copies of each of the four canonical histones, H2A, H2B, H3 and H4 (figure 1), which themselves are the most abundant DNA-binding proteins in the nucleus. These structures reveal in
exquisite detail the molecular interactions between the histones and the DNA. In particular, the DNA-binding surface of the octamer forms a ramp into which arginine residues protrude at 3.3–3.4 nm intervals corresponding to the length of a single turn of the DNA double helix. Although structures of the NCP determined by crystallization with different DNA sequences (often rather small differences) exhibit minor variations in the structural details depending on the DNA sequence, these variations can all be accommodated in the framework of a single overall structure [6]. But, how relevant is this structure to the biological function of the NCP? When free in solution, the NCP is a highly dynamic particle, a property that is intimately associated with its biological role in the regulation of chromatin structure and gene expression. For example, in solution, the extreme ends of the DNA segment bound to the octamer rapidly unwrap and rewrap [7–10], although there is evidence that the nucleosome can adopt a different (undefined) structure in solution [11]. Because in the crystal the DNA wrapped by the octamer is relatively static, the nature of this dynamic structural variation is not apparent in the crystal structures. Equally pertinently, crystallization of NCPs has so far only been possible using a very narrow range of DNA lengths (145–147 bp), yet the basic N-terminal regions of certain histones are known to bind to DNA external to this length in what is termed ‘linker’ DNA. These interactions may contribute to the topological constraint of DNA by the histone octamer, but, for technical reasons, are not seen in the crystal structures of NCPs [12].

Although there may be caveats with regard to the relevance of the NCP crystal structure in certain biological situations, the details of how the DNA is wrapped around the octamer are fundamental to the understanding of how nucleosomes are assembled, how particular DNA sequences favour nucleosome formation and how nucleosome arrays are organized. Two characteristics of histone–octamer-bound DNA are notable—it is tightly bent and is wound in a negative supercoil. In solution, the DNA polymer is stiff with a persistence length of approximately 50 nm corresponding to approximately 150 bp or 14.3 double-helical turns of B-DNA [13]. This is the length over which, in solution, its behaviour approximates to that of a straight rod. Yet, in the NCP, the same length of DNA is coiled with an average radius of curvature of 9 nm. The widths of the DNA grooves, both major and minor, are consequently greater on the outside of the bent DNA than on the inside, a difference that must be accommodated by the averaged conformation of the individual base steps. A particular limiting parameter is the width of the minor groove, which on the inside of the wrapped DNA can approach 3 Å, much lower than the average solution value of approximately 8 Å [14] (figure 2). This imposes strong steric constraints on the DNA because narrowing of the minor groove is impeded by the presence of the exocyclic 2-amino group of guanine [16,17]. This steric constraint is consistent with the preferred sequence organization of octamer-bound sequences in which AA/TT dinucleotides repeating every 10 bp (with GC dinucleotides 5 bp out of phase) and positioned where the minor groove contacts the octamer impose a curvature favourable to nucleosome formation and stability [18,19]. The minor groove contacts are stabilized by the phased arginine residues extending into the DNA-binding ramp of the octamer [2,14]. Within the octamer-bound DNA, this type of sequence organization—even if present in only a limited region (80 bp or less [20,21])—can function as a nucleosome-positioning sequence and locate the octamer very precisely relative to the DNA sequence.
The preferred sequence pattern with A/T- and G/C-containing dinucleotides placed where the minor groove is, respectively, on the inside and outside of tightly bent DNA is also characteristic of other DNA–protein complexes, including those containing transcription factors such as the factor for inversion stimulation (FIS) and the adenosine 3′,5′-cyclic monophosphate receptor protein (CRP) and certain structural domains of type II topoisomerases [22–24]. In these complexes, the diameter of the protein spool, like that of the bacterial histone-like protein from *Escherichia coli* strain U93 (HU) protein, is comparable with that of the histone octamer [25–27]. The orientational preferences of the DNA sequence in such complexes must thus represent an intrinsic property—bending anisotropy—of the DNA itself. Experimental and theoretical aspects of DNA bending are discussed in more detail in appendices A and B, respectively.

2. DNA-bending anisotropy

Sequence-dependent bending anisotropy (appendix C; preferable to the misnomer ‘anisotropic flexibility’ [18]) is a fundamental physico-chemical property of DNA that structurally distinguishes one region of DNA from another and thus constitutes an additional form of encoded information. The ability to bend preferentially in a particular direction is normally apparent over sequence ranges of 20–150bp, although, inevitably, there are some exceptions. For example, the naturally occurring circle-forming kinetoplast DNA molecules of *Crithidia* can extend for at least 200 bp [28,29].
For a given sequence, a preferred DNA-bending direction in a DNA molecule in solution is conferred by the sum of preferred conformations of individual base-steps such that the range of conformations potentially adopted by such base steps is limited (reviewed in Travers & Thompson [30]). In other words, these base steps are conformationally rigid relative to other base steps. The adopted resulting trajectory (configuration) of the DNA double helix is then, relative to isotropic bending, restricted. Anisotropy, by definition, means the restriction of some degrees of bending freedom and so results in the occupation of a smaller volume of configurational space relative to a chain that bends isotropically (figure 3). In this sense, the properties of the DNA molecule can be described by the Boltzmann formulation for entropy (appendix D).

In the context of DNA–protein interactions, this implies that the closer the trajectory conferred by a DNA sequence corresponds to the path of the DNA in a DNA–protein complex, the smaller will be the entropic penalty on binding. In other words, possibly paradoxically, a DNA sequence preferentially trapped by a DNA-bending protein could be stiffer than average. Nevertheless, this will be true only if the net enthalpic cost of distorting the DNA structure on binding is small (appendix E). The energy required for bending a DNA molecule increases substantially with curvature. For DNA, in solution, the cost of tight bending is high, such that the formation of DNA circles with the same radius of curvature as DNA in the nucleosome normally requires the participation of other DNA-bending proteins. However, the charge neutralization on one face of the DNA polymer on protein binding facilitates bending because the mutual charge repulsion on the remaining charged surface on the other face of the DNA (to become the outside of the bend) itself induces bending [31,32] (figure 4). Similar considerations may apply in solution, particularly in the presence of multi-valent cations [33]. Indeed, Hud & Plavec [34] have attributed anisotropic DNA sequence-dependent curvature to the experimentally observed preferential sequence-dependent build-up of cations in the minor grooves of AT-rich sequences, and the major grooves of GC-rich sequences, resulting in a charge imbalance of opposing faces of the double helix.

The tight bending of DNA by proteins normally requires the deformation of base-step conformations beyond those sampled in solution as a result of thermal fluctuations. The nature of the base-step deformations induced by

Figure 3. Cartoon of the limitation of occupation of configurational space conferred by (b) bending anisotropy contrasted with (a) isotropic bending.
DNA-bending proteins depends strongly on the precise nature of the interaction. Some proteins, for example, high mobility group B (HMGB) proteins, TATA-binding proteins, the bacterial chromosomal proteins HU and integration host factor as well as the _lac_ repressor, widen the minor groove by intercalating one or more amino acids between base pairs [15,35–38]. Others, such as FIS and CRP [22,23], bind in the major groove and concomitantly widen the minor groove. In contrast, charge neutralization by the histone octamer takes the form of the insertion of an arginine residue deep into the minor groove at approximately 10 bp intervals [2,14]. The high positive charge density on the arginine facilitates an extreme narrowing of the minor groove comparable to that observed in the C-form of DNA [39,40]. Concomitant with this narrowing is a change in the structure of the phosphate backbone favouring the BII rather than the BI conformation [41]. The propensity to adopt the BII conformation is sequence dependent such that, in solution, it is freely adopted by G/C-containing, but not A/T-containing, steps [41]. This sequence preference is apparent in the formation of extreme distortions, in the form of DNA kinks into the minor groove [4,42]. These kinks occur principally at CA/TG and CC/GG steps but not, as originally predicted, at TA steps [5].
To what extent are these considerations consistent with experimental observations? The selection of DNA sequences for binding to the histone octamer in vitro yielded a set of sequences with the highest determined affinity for the octamer [43,44]. Notable among these is the 601 sequence that positions the octamer very precisely and which also possesses a high bending anisotropy, apparent by direct visualization using atomic force microscopy, using gel retardation assay and also its capacity for facile circle formation using nucleosome length DNA fragments. Like other nucleosomal DNA sequences, the 601 sequence has short A/T blocks repeating every approximately 10 bp interspersed with short G/C blocks in the opposite helical phase. Notably, the 601 DNA sequence has a higher overall stacking energy (stacking energy is the attractive energy between successive base pairs in the DNA double helix, and, to a first approximation, correlates with DNA stiffness) than other sequences that bind the octamer with lower affinity [45]. The 601 sequence is thus an example of a DNA sequence whose bending anisotropy conforms to the DNA-binding region of a particular protein [46]. In this situation, the entropic penalty for binding will be low relative to an isotropically bendable DNA and any further distortion of DNA structure necessary to achieve tight binding will be energetically less costly than that of entropy. A similar phenomenon is also apparent in the cleavage of DNA by the nuclease DNase I. This protein binds in the minor groove of DNA and bends DNA away from the protein surface. It thus preferentially cleaves sequences with a wide minor groove. Among the sequences with the highest rate of cleavage are some conformationally rigid trinucleotides (e.g. GGC/GCC) that naturally adopt a wide minor groove [47]. Again, the free energy required to adopt the preferred binding conformation is minimized.

These considerations bear on the issue of DNA flexibility and its theoretical treatment. Classically, the behaviour of a DNA molecule in solution has been described with substantial success by the worm-like chain (WLC) model. A major assumption of this model is that the polymer can bend isotropically. While sequence heterogeneity ensures that, on average, this is true for most DNA molecules longer than approximately 200 bp, for short DNA sequences with strong bending anisotropy, the applicability of the WLC model is less certain. Experimentally, the axial flexibility of DNA molecules has been determined by the methods measuring the rate of cyclization of short DNA molecules [48,49], assuming that flexibility is directly related to this rate. In essence, this method is conceptually analogous to the binding of DNA by the histone octamer because both require that the DNA adopt a particular trajectory—a circle for flexibility measurements and a low pitch helix for the octamer. In the circularization assay, the 601 positioning sequence cyclizes rapidly relative to other sequences, but the conclusion that this sequence is more axially flexible than others is possibly misleading. The high relative rate of circularization could result simply from the anisotropic bending placing the two ends of a 601 DNA fragment of appropriate length in close spatial proximity. Any further bending required for forming a circle would then be small. More accurately, in these circumstances, the calculated ‘flexibility’ should be termed ‘apparent flexibility’ [50]. Again, because in a circularization assay, determinations of axial and torsional flexibility are not independent, conclusions suggesting that a strongly axially anisotropic DNA sequence is also correspondingly torsionally flexible may be misleading.
Of particular interest in this context are certain abundant DNA architectural proteins, HU and its paralogues in bacteria and the HMGB proteins in eukaryotes, which bind DNA with little or no sequence selectivity and enhance both the apparent axial and torsional flexibility as determined by cyclization kinetics. They are biologically important because they facilitate the formation of tightly bent DNA loops in protein–DNA complexes involved both in recombination and also in transcriptional regulation [51–54]. These proteins, although structurally disparate in bacteria and eukaryotes, both bend and untwist DNA, and are characterized by rapid on/off rates for DNA binding. The imposed transient bend(s) coupled with the changes in DNA twist allow sequences that are distant from each other on the physical map to be brought into an appropriate close spatial proximity. In effect, these proteins by imposing a bend restrict the configurational space occupied by a DNA molecule but because of the fast on/off rates enable rapid sampling of different configurations. The net effect of these proteins on DNA is thus to reduce the apparent persistence length [55,56].

Although bending anisotropy is an important determinant of selectivity in the formation of protein–DNA complexes with tightly bent DNA, the axial flexibility of the molecule remains an important parameter. In particular, the bending anisotropy conferred by a particular DNA sequence organization is dependent not only on base-stacking interactions, but also on the structure of the ordered water envelope surrounding the double helix [57]. This water structure is strongly temperature dependent over the physiological range (0–40°C), being disrupted at higher temperatures with a corresponding abrogation of bending anisotropy [58,59]. One consequence of this variation is that, at higher temperatures, DNA sequences become more equivalent in terms of the occupation of configurational space and consequently more equivalent in terms of the entropic penalty for adopting a particular trajectory. In other words, the loss of bending anisotropy in principle increases axial flexibility. This means that the balance between the entropic and enthalpic components for the selection of DNA sequences by, for example, the histone octamer would be temperature dependent. At low temperatures, the entropic component is more important; at higher temperatures, the enthalpic component assumes greater prominence. Thus, in contrast to the 601 sequence, which was selected at 4°C, another selection at higher temperatures (37°C/25°C) resulted in the isolation of high-affinity sequences of low stacking energy (and hence more flexible) [60]. Among these were repeating sequences of the form (CAG)$_n$, which lack the sequence patterns for anisotropic bending and accordingly do not position nucleosomes precisely [61], again contrasting with the sequences selected at 4°C. The inference is that the relative affinity of different DNA sequences for the histone octamer is not invariant, but depends on the assay conditions, as has subsequently been demonstrated experimentally [45].

The relative loss of bending anisotropy at higher temperatures and the consequent acquisition of a more uniform trajectory among DNA sequences implies a loss of information content. This would be consistent with the differences in average positioning precision of octamer-binding sequences selected at 4°C and at 37°C/25°C, and raises the question as to what is the selection balance for nucleosome formation in vivo and to what extent is this balance influenced by chromatin-remodelling assemblies that shift the positions of nucleosomes?
3. The role of DNA supercoiling

In most organisms, the DNA is negatively supercoiled. This both facilitates the packaging of DNA in coiled structures and, at least in bacteria, enables processes such as transcription initiation, recombination and the initiation of DNA replication, which involve the unwinding of the DNA double helix. But, the supercoiling of DNA also has important implications for DNA–protein recognition. The energy required to supercoil DNA increases the motions of the polymer as well as influencing the available conformations of individual base steps. At high superhelical densities, the DNA molecule, depending on the precise sequence, can locally assume different secondary structures. These would include, among others, the regions of local DNA strand separation [62], cruciform structures [63,64] and slipped loops [65,66] (appendix F). The result is that a supercoiled molecule would sample an increased range of configurations and base-step conformations and thus for many DNA-binding proteins, the entropy penalty for complex formation would be increased.

However, DNA superhelicity is distributed in the molecule in two forms: a change in twist, manifest as unwinding in negative superhelicity, and a change in the configuration of the molecule such that the torsion is expressed in a coiling of the double-helical axis (writhe). For negatively supercoiled DNA, this coiling can be left- or right-handed depending on whether the overall form of the DNA molecule is toroidal or plectonemic, respectively (figure 5). Because many of the abundant DNA-binding proteins that organize eukaryotic or bacterial chromatin bind negatively supercoiled coils, supercoiling facilitates their binding and can enhance the assembly of protein–DNA complexes such as the NCP. These proteins also play a role in maintaining supercoiling levels. For example, in *Escherichia coli*, HU, which locally constrains left-handed DNA toroids [25,26], decreases gene...
activity in chromosomal regions with a high supercoiling potential (as inferred from DNA gyrase-binding sites) and conversely increases activity in the regions with a lower supercoiling potential [68]. This mode of action could thus be the molecular equivalent of a mechanical governor [69] (figure 6).

More importantly, negative supercoiling can drive the formation of functional contacts that would otherwise occur infrequently. A good example of this phenomenon is the binding of the bacterial RNA polymerase to DNA sequences about 100 bp in front of certain promoters. In such cases, the rate of contact formation can be enhanced by up to 100-fold by negative superhelicity [70]. The DNA sequence of such regions indicates that they have the potential to bend anisotropically to form a left-handed coil [71]. The enhanced contacts thus suggest that negative superhelicity shifts the sampled DNA configurations towards a range that is more appropriate for binding to the polymerase.

One implication of this and related findings is that DNA supercoiling can fine-tune the path of the DNA double helix so as to optimize DNA–protein contacts. In a circular DNA molecule (or a linear constrained topological domain), the extent and nature of the coiling induced by negative DNA supercoiling is dependent on both the superhelical density (the higher the density, the greater the coiling and hence a lower pitch angle; the inclination of the coil to the axis at right angles to the superhelical axis) and the DNA sequence. For the latter, this is because the structural periodicity in DNA coils is dependent on the sense of coiling: for left-handed toroidal coils, the structural periodicity is $<$10.5 bp (the periodicity of relaxed DNA) and for right-handed coils, this periodicity is $>$10.5 bp. The structural periodicity can be encoded in the DNA sequence as a sequence periodicity and thus the DNA sequence can potentially influence the form of coiling in a supercoiled DNA molecule.

Phil. Trans. R. Soc. A (2012)
By enhancing the occurrence of preferred DNA configurations for protein binding, DNA supercoiling thus can potentially act as an organizer of the chromatin structure. More importantly, because in bacteria, the superhelical density of DNA varies depending on energy availability and different architectural proteins bind optimally to DNA of different superhelical densities, changing the superhelical density of DNA could change the overall organization of chromatin with consequent effects on gene expression. In contrast to the discontinuous (on/off) mode normally envisaged for the interaction of transcription factors with their target sites, superhelical density effectively acts as an analogue regulatory mode [72] specifying the global expression pattern of the digital linear code [73]. By analogy to electronics where the analogue signal (e.g. current) is converted into a digital number proportional to the magnitude of the current, in the case of the DNA, we assume that the analogue signal (genome-wide supercoil dynamics) is converted into the digital gene expression pattern corresponding to the magnitude of superhelical density. This dual informational content of the DNA molecule and the inherent convertibility of the digital and analogue codes thus fully conforms to the original postulate of Schrödinger [74] of an organizational principle for generating ‘order from order’ as a major distinction between the inanimate and living matter.

G.M. thanks Marc-Thorsten Hütt for inspiring conversations. For appendix B, we are indebted to David Swigon, Wilma Olson and Irwin Tobias.

Appendix A. Base-step conformational variation and DNA stiffness

The stiffness of a DNA molecule is strongly dependent on the nature of the DNA sequence, but what are the sequence parameters that determine physico-chemical properties? The fundamental observation is that, for a given DNA sequence, with the exception of those that confer intrinsic curvature, when the average number of hydrogen bonds/base pair is varied between two and three by base-analogue substitution, there is a direct correlation between this quantity and the persistence length [61]. This implies that one major factor determining the persistence length is the average base-pair rigidity, with G–C base pairs being more rigid than A–T base pairs. Similarly, when an intrinsic curvature is conferred by phased oligo(dA:dT) tracts, the presence of bifurcated hydrogen bonds between adjacent A–T base pairs may also act to enhance the bending stiffness and anisotropy [75,76].

Locally, the quantitative effect of the number of hydrogen bonds per base pair on DNA stiffness depends on sequence context and, in particular, on the types of base steps. Of the three types of base step, pyrimidine–purine (YR), purine–purine/pyrimidine–pyrimidine (RR/YY) and purine–pyrimidine (RY), the YR steps are the least, and the RY steps the most, thermally stable (table 1). Within each class of base step, the experimentally determined stacking energy—and hence the stability—is related to base composition such that steps containing only A–T base pairs have on average both a lower stacking energy and melting energy than those containing only G–C base pairs [77] (table 1).

The persistence length of DNA is, in essence, a measure of the average magnitude of thermally induced variations in the axial direction. In relating variations in the conformation of individual base steps to those of the local
Table 1. The properties of the 10 base steps in free DNA. The base steps are grouped according to the type and base composition. Beneath each base step are listed successively the stacking energy, the melting energy (in red) (both in kJ mol\(^{-1}\) from [77]), the range of conformations adopted by the base step from El Hassan & Calladine [78] as shown in figure 7, the twist, roll and X-disp (base-pair displacement) index (a measure of the percentage probability of one of two phosphates in the base step assuming a BII conformation) [41], and the preferred minor groove orientation of the step in chicken nucleosomal DNA, \(I = \text{in}, O = \text{out}, i = \text{in as a minor preference}, \text{np} = \text{no preference} [18, 79]\). The terms ‘flexible’ and ‘rigid’ beneath the YR and RY steps are taken from [80]). Note that the base steps that are most deformable when bound to a protein are AA/TT, AT, TA, CA/TG and CG [81]. Estimates of the stacking and melting energies of base steps were determined from a single sequence context.

<table>
<thead>
<tr>
<th></th>
<th>narrow minor groove</th>
<th></th>
<th>wide minor groove</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>flexible</td>
<td>stiff</td>
<td></td>
</tr>
<tr>
<td>YR</td>
<td>TA</td>
<td>CA/TG</td>
<td>CG</td>
</tr>
<tr>
<td></td>
<td>–0.79</td>
<td>–2.30</td>
<td>–3.81</td>
</tr>
<tr>
<td></td>
<td>–0.50</td>
<td>–3.26</td>
<td>–6.02</td>
</tr>
<tr>
<td></td>
<td>extensive</td>
<td>extensive</td>
<td>bimodal</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>42</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>i/O</td>
<td>np</td>
</tr>
<tr>
<td>RR/YY</td>
<td>AA/TT</td>
<td>AG/CT</td>
<td>GA/TC</td>
</tr>
<tr>
<td></td>
<td>–4.64</td>
<td>–4.44</td>
<td>–5.98</td>
</tr>
<tr>
<td></td>
<td>–4.35</td>
<td>–5.40</td>
<td>–6.95</td>
</tr>
<tr>
<td></td>
<td>restricted</td>
<td>nd</td>
<td>restricted</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>np</td>
<td>np</td>
</tr>
<tr>
<td>RY</td>
<td>AT</td>
<td>AC/GT</td>
<td>GC</td>
</tr>
<tr>
<td></td>
<td>–5.61</td>
<td>–7.57</td>
<td>–9.07</td>
</tr>
<tr>
<td></td>
<td>–5.31</td>
<td>–8.53</td>
<td>–11.3</td>
</tr>
<tr>
<td></td>
<td>restricted</td>
<td>intermediate</td>
<td>bimodal</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>np</td>
<td>O</td>
</tr>
</tbody>
</table>

Configuration of the DNA chain, an important consideration is the extent of the conformations that can be assumed by individual base steps. Examination of the conformations of base steps determined in crystal structures reveals that whereas the possible conformations of YR base steps span a wide range, and hence are termed ‘flexible’, those of RY steps are more restricted [78, 80] (figure 7). A complication is that analysis of DNA crystals, including both A- and B-form structures, suggests that the distribution of conformations adopted by base steps containing only G–C base pairs, i.e. CG, GG/CC and GC, has a
Figure 7. Different base steps can adopt varying degrees of conformational flexibility [78]. The figure plots the data from DNA crystal structures showing the variation of the base-step parameter slide (which determines the displacement of the base-pair axis from the double-helical axis) with the roll and twist parameters. Negative slide, positive roll and low twist favour a wide minor groove, while positive slide, negative roll and high twist favour a narrow minor groove. The continuous ranges and bimodality of conformations are shown. For reference, the average roll angle values for outward and inward phasing minor grooves for DNA wrapped on a nucleosome are approximately $10^\circ$ and approximately $-7^\circ$, respectively [5,42]. Reproduced with permission.

bimodal character—in other words, each of these base steps can adopt two distinct conformations, but both of them occupy a restricted range. This bimodality is not apparent when the analysis is confined to B-form DNA crystals [80].

When DNA bends, while the intrinsic stiffness of a base step may determine the variation in bending, the geometry of any preferred conformations will determine the direction of bending, i.e. towards the minor or major groove. For base steps, including G–C base pairs, important factors include the presence of the 2-amino group (of guanine) in the minor groove, the pronounced dipole moment associated with the base pair and the ability to adopt both BI and BII phosphate conformations, correlating, respectively, with wide and narrower minor grooves [41]. The 2-amino group limits compression of the minor groove while in base steps containing only G–C base pairs, the dipole moment by shifting the relative alignments of successive base pair results in the displacement of the central axis of the base pair away from the double-helical axis. The preference for a wider minor groove, coupled with such displacement, confer characteristics more typical of A-form DNA. In contrast, A–T base pairs lack both a strong dipole moment and a 2-amino group on the purine so that base steps containing only A–T base pairs, e.g. AA/TT and AT, stack so that the central axis of the base pair is not displaced from the double-helical axis.

When bending is increased on the surface of the histone octamer, the conformational restrictions of base steps apparent in free DNA become more restrictive because of the changed groove width dimensions. Where the minor
groove is wide on the outside of the bent DNA, the DNA structure approaches that of A-DNA, whereas on the inside, the structure more closely resembles that of C-DNA (an over-twisted variant of B-DNA) [40]. In this situation, the selected sequence patterns depend both on the precise sequence and base composition of the bound DNA. For a DNA of average base composition, the finding of Satchwell et al. [18] for chicken erythrocyte nucleosomal DNA that A/T-rich and G/C-rich short sequences preferentially occupy positions where the minor groove is, respectively, narrow and wide correlates well with the known properties of the base steps. Note that the periodicity of the preferred occurrences of such sequences likely results from the preferential exclusion of G/C-rich sequences where the minor groove is narrowed and conversely for A/T-rich sequences. So far, only two examples of bimodality (where a base step preferentially occurs at both wide and narrow minor grooves, but to a lesser extent in between) have been noted. These are CA/TG (a flexible base step in free DNA) in chicken erythrocyte nucleosomal DNA [18] and GG/CC (a bimodal step in crystal structures) in nucleosomes reconstituted on ovine DNA in vitro [79]. Importantly, the properties of individual base steps are influenced by their sequence context such that the preferred pattern of dinucleotides (base steps) in nucleosomal DNA reflects that of tri- and tetranucleotides [18,41,82]). For example, the periodic pattern of AA/TT occurrences in nucleosomal DNA follows that of AAA/TTT occurrences, while the periodic pattern of GC occurrences follows more closely that of GGC/GCC occurrences and not that of, for example, TGC/GCA occurrences [18]. This is likely because, for example for AA/TT, while the base step has a preferred stacking conformation in isolation, in a sequence context such as CAAG, the flanking base pairs, which also stack on their adjacent A–T base pairs, will affect the stability of the AA/TT stacking.

In summary, the elastic properties of DNA are likely determined by two principal factors: first, the stiffness of the individual base pairs that, in turn, is reflected in the stiffness of the base steps and, second, the type of base step (YR, RR/YY, RY) substantially affects the extent and stability of different conformations that the base step can adopt.

Appendix B. Computer modelling of sequence dependencies

There has been much activity and progress in recent years on discrete computer models that have been developed to embrace the sequence-dependent elasticity of DNA in a way that closely resembles the detailed DNA structure. The most common of these discrete models treat DNA as a collection of rigid subunits representing the base pairs. This description has long been used by chemists to characterize DNA crystal structures [83,84]. For each base pair, the DNA configuration is specified by giving its location in space and its orientation measured by an embedded orthonormal frame. In particular, the relative orientation and position of a base pair and its predecessor are specified by six kinematical variables termed, respectively, tilt, roll, twist, shift, slide and rise. In the simplest (notably the so-called dinucleotide models), the elastic energy is taken to be the sum of the base-pair step energies, each of which is a quadratic function of the kinematical variables.
Some of the most detailed information about DNA structure and flexibility come from the analysis of X-ray crystal structures and nuclear magnetic resonance spectroscopy experiments. Empirical estimates of intrinsic values and elastic moduli have been deduced from the averages and fluctuations of base-pair step parameters in high-resolution DNA protein complexes [80]. Here, departures from ideal behaviour found by Olson and collaborators include intrinsic bending (in the roll variable), bending anisotropy, inhomogeneity in twisting to bending stiffness ratio, twist–roll coupling, twist–stretch coupling and shearing anisotropy. Some other experimental methods used to examine elastic behaviour of longer segments in which the effects of individual base-pair steps are spatially averaged are as follows: cyclization [85,86]; fluorescence resonance energy transfer [87]; gel mobility [88]; single-molecule stretching [89,90], and twisting [91–93]. The sequence-dependent nature of DNA deformability has been independently confirmed by research aiming to deduce DNA elastic properties from molecular dynamics simulations [94,95].

For the afore-mentioned dinucleotide model with quadratic energy, variational equations have been derived by Coleman et al. [96] and equilibrium configurations for plasmids of various compositions and end conditions have been found [96,97]. These later configurations include: (i) multiple equilibria of ligand-free DNA o-rings (plasmids that are circular when stress free), (ii) minimum configuration of DNA o-rings with bound intercalating agents, (iii) optimal distribution of intercalating agents that minimizes elastic energy of DNA o-rings, (iv) collapsed configurations of DNA o-rings subject to local overtwisting, (v) minimum energy configurations of intrinsically straight DNA plasmids with various distributions of twist–roll coupling, and (vi) minimum energy of S-shaped DNA subject to local overtwisting. In an unpublished work, Kocsis & Swigon have analysed the detailed mechanical response of the calibrated discrete model to stretching and twisting and found that bifurcations give rise to multiple equilibria for given extension or applied force and shearing instability. These authors also found that stretching gives rise to overtwisting of DNA, in accord with experimental results [98,99].

Theoretical work has been extended to account for electrostatic repulsion and thermal fluctuations and applied to the study of minimum energy configurations and looping-free energies of LacR-mediated DNA loops [100], and minimum energy configurations of free segments of promoter DNA bound to class I and class II CRP-dependent transcription activation complexes [101]. Theory has also been combined with hydrodynamic interactions of the solvent; simulations of the model using immersed boundary method have been used to study the dynamics of DNA supercoiling [102].

There have been suggestions that the local energy of DNA deformations may depend on the composition, or even the deformation, of more than just the immediate base-pair neighbours. For example, trinucleotide and tetranucleotide models have been proposed to account for some DNA structural features [82], and they also seem to better represent averaged DNA properties extracted from molecular dynamics simulations [94,95]; the mechanical theory of such models has not yet been constructed. Extensive stretching or twisting can induce the transition of DNA to alternative conformations with disrupted base pairing, but it is not known whether such conformations play any biological role. DNA kinking—a higher order response to bending associated with disruption...
of base stacking—has been studied recently [103–107]. This work can suggest an explanation for unusually large cyclization probabilities of certain special DNA sequences [108,109], but see remarks about the latter in Du et al. [110]. Nevertheless, recent experimental visualization of 94bp DNA minicircles show no sharp kinks [111].

A new coarse-grained model of DNA was recently proposed in which each base pair is represented by two beads and energy is prescribed for interactions between a bead in one strand and 11 consecutive beads in the opposite strand [112], which appears to better describe the dependence of persistence length on the concentration of counterions.

Olson and co-workers have introduced protein-bound segments identical to those found in high-resolution structures in computer simulations of DNA loop formation [100,113] and ring closure [114,115]. In the work on ring closure, the properties of DNA chains decorated with the non-specific *Escherichia coli* histone-like architectural protein HU are particularly interesting. The extreme deformations of DNA associated with the binding (significant bending and untwisting of successive base pairs accompanied by large dislocation of the helical axis) account for the observed cyclization propensities of minicircles formed in the presence of HU [115]. Here, despite the low levels of protein (one HU dimer per 100–1500bp), most successfully closed molecules contain two or more bound proteins [114]. That is, the protein binds preferentially to closed DNA as opposed to linear DNA, while, additionally, the build-up of protein depends on the chain length. The duplexes with fewer bound HU dimers are torsionally relaxed, and the minicircles 5–6 bp shorter or longer with more bound proteins are negatively supercoiled.

As is evident from the elongated forms of simulated 200bp minicircles, the binding to circular DNA is associated with a straightening of the intervening protein-free segments. In other words, the added proteins counterbalance the bending stress on cyclic DNA and thereby reduce the total free energy of the system. Furthermore, in contrast to bare DNA, where base pairs must twist in order to bring the ends of the duplex in register for ligation, duplexes with a non-integral helical repeat simply bind additional HU upon cyclization.

To quantify the change in twist associated with each protein-binding event, Britton et al. [116] introduced a new quantity called the *twist of supercoiling*, different from the step parameter, called the *Twist*, which is one of the six rigid-body parameters used to describe the spatial arrangement of successive base pairs [117]. Unlike the step parameter, Twist gives an integral linking number when summed and combined with the writhing number of a closed DNA. HU-binding events like those introduced in the calculations lower the twist of supercoiling by approximately 45° compared with canonical B-DNA, depending upon the choice of high-resolution structure.

Effects of local sequence-dependent bending anisotropy of base-pair steps on DNA cyclization propensities have also been examined [118,119]. Several well-known structural features of DNA (including the presence of intrinsic curvature, roll–twist coupling, or enhanced pyrimidine–purine deformability) enhance the computed cyclization propensities. Moreover, periodically distributed roll–twist coupling reduces the magnitude of oscillations in the Jacobson and Stockmayer [120] cyclization factor *J* (seen in plots of *J* versus chain length) to the extent found experimentally.
Appendix C. Stiffness of elastic rods

(a) Axial stiffness

An elastic rod of length \( L \) subjected to a tension, \( T \), will exhibit an axial extension \( e \). In continuum mechanics, the stiffness, \( T/e \), will be \( EA/L \), where \( E \) is Young’s modulus of the material and \( A \) is the cross-sectional area. The stored strain energy, \( U \), equal to the work done in producing the extension is \( 1/2Te = 1/2(EA/L)e^2 \).

(b) Torsional stiffness

A solid circular elastic rod of length \( L \) subjected to equal and opposite end twisting moments, \( N \), will exhibit a relative rotation of \( \theta \) at its ends. The torsional stiffness, \( N/\theta \), is given by \( GJ/L \), where \( G \) is the shear modulus of the material and \( J \) is the polar moment of inertia of the circular cross-sectional area. Here, \( U = 1/2(GJ/L)\theta^2 \).

(c) Bending stiffness

(i) The isotropic case

A rod of circular cross section has equal bending stiffness in every direction. If bent by equal and opposite end moments, \( M \), the rod deforms into a planar circular arc with a radius of curvature \( R \). The reciprocal of \( R \) is the curvature, \( \kappa \), and the bending stiffness \( M/\kappa \) is given by \( EI \), where \( E \) is Young’s modulus of the material and \( I \) is the moment of inertia of the cross-sectional area about the bending axis. Here, the stored energy is \( U = 1/2EI\kappa^2 \).

(ii) The non-isotropic case

A rod with a non-symmetric cross section (think of a ruler) has two principal axes where the bending stiffness takes its maximum and minimum values. Bent about either of these special axes, the rod will deform into a circular arc, and the bending stiffnesses can be written as \( (M/\kappa)_1 = EI_1 \) and \( (M/\kappa)_2 = EI_2 \). Here, \( I_1 \) and \( I_2 \) are the principal moments of inertia of the cross-sectional area. Bending about other axes can be studied using the principle of superposition.

(iii) Change in shape and bending stiffness of DNA

Sequence-dependent structural changes of DNA might have a number of bending effects (alone or in combination) that need to be distinguished.

First, the changes might give the molecule an initial curvature, \( \kappa_0 \), in its natural unloaded state, leaving it (say) with its original isotropic bending stiffness \( EI \). In this case, under a bending moment that further increases the curvature to its total value, \( \kappa_T \), we can write \( M = EI(\kappa_T - \kappa_0) \), with \( U = 1/2EI(\kappa_T - \kappa_0)^2 \).

Second, the changes might leave the bending isotropic, but increase or decrease the isotropic stiffness \( EI \) over a particular region. This could be achieved in a metal rod, for example, without changing the cross-sectional area (and therefore the weight per unit length) if over a region the solid circular cross section was replaced by a hollow tube with its greater \( EI \).

Phil. Trans. R. Soc. A (2012)
Third, the sequence changes might destroy the isotropy over a linear region of
the molecule replacing the original $EI$ by two principal stiffnesses $(EI)_1 > (EI)_2$. It might be tempting to suppose that $(EI)_1 > EI > (EI)_2$, but there is really no
guarantee that this would be the case.

(iv) Worm-like chain model

The WLC model in polymer physics, as originally formulated by Kratky and
Porod, is used to describe the behaviour of a semi-flexible polymer, such as DNA. It usually considers an isotropic rod that is continuously flexible in contrast to a jointed chain that is only flexible between segments. It is used to model the behaviour of a polymer under thermal agitation, drawing on Boltzmann’s formulations.

(v) Mechanical modelling of DNA

The various mathematical approaches to the mechanics of DNA, including atomic modelling, base-pair modelling, continuum modelling and finite-element analysis, are reviewed by Travers & Thompson [30].

Appendix D. Entropy and the second law of thermodynamics

Entropy can be thought of as a measure of the ‘randomness’ of a system, and the second law of thermodynamics states that it will have a larger (strictly, not smaller) value after some process takes place inside the system.

It was the brilliant Austrian physicist Ludwig Boltzmann who, in 1877, gave us the modern definition of entropy. This definition, for a classical system of $N$ featureless particles (simple molecules in a gas, say), relates to the $6N$-dimensional phase space spanned by the positions and momenta of all the particles. This phase space is given a coarse graining, by imagining it divided into a number of subregions that are conventionally called ‘boxes’.

Ensembles of points representing states of the system that are macroscopically indistinguishable from one another are considered to be grouped together in the same box. Correspondingly, points belonging to different boxes are deemed to be distinguishable on a laboratory scale.

The Boltzmann entropy, $S$, for a state represented by a point of the phase space is now defined as

$$S = k \log V,$$

where $V$ is the volume of the box within which the point lies. Here, the Boltzmann constant, $k$, has the value \(1.38 \times 10^{-23}\) JK$^{-1}$.

Appendix E. Barriers of energy, enthalpy and entropy

In discussing the ability of a DNA molecule to bend, bind and generally perform its various tasks, it will be necessary to consider the changes in the thermodynamic variables, internal energy, enthalpy and entropy. In particular, we shall speak about these in terms of perceived barriers, costs or penalties that might inhibit the desired outcome.
The simplest of these variables is the *internal energy*, which corresponds to the elastic strain energy described in connection with the distortion of elastic rods. To get to a new deflected state of higher internal energy, the work that must be done represents a well-defined barrier or cost.

Next, we must consider the *enthalpy*, \( H \), which is just the internal energy, \( U \), of a thermodynamic system with the addition of pressure times volume, \( pV \). This latter term takes account of any work done against the environment. So, in instances where there is a significant interaction with the local environment, it is more appropriate to discuss barriers and costs in terms of enthalpy rather than internal energy.

The final barrier or cost is that of *entropy*, which we have formulated earlier as \( S = k \log V \). When we perceive that a desired configuration of the molecule requires a change from a rather random or natural state to a highly unusual or organized state, it is appropriate to speak of an entropic barrier or cost. The implication is that the organized state will not be easily accessed under the action of random thermal agitation.

### Appendix F. Twisting, linking and writhing

The double helix of most DNA is *right handed*, just like right-handed screws on sale in the DIY shop: this includes the well-known A and B forms, while the Z form is by contrast *left handed*. To transcribe the genetic code, DNA must screw through an RNA polymerase. This rotation, which can be at about 10 turns per second, can induce large twisting stresses in the DNA. If, under stress, the intrinsic helical twist of the molecule is increased, we say the molecule is *overwound*. Conversely, if this twist is decreased, we say it is *underwound*.

Now when an initially straight elastic rod is highly twisted, it is known to writh into a three-dimensional helical configuration, as illustrated in figure 8, where we see a pulled and twisted rod deforming first into a writhed helical state, and then starting to form a self-contacting ply. For more details, see van der Heijden *et al.* [121].

This transformation of twist into writhe alleviates the twisting energy at the (lower) expense of the bending energy. In DNA, this writhing produces what are called supercoils, and two general varieties have been discussed in the main text. Here, we shall focus mainly on the *interwound* type, otherwise known as a *ply*, for which the following rules apply:

- positive (right-handed) twist \( \rightarrow \) left-handed ply with positive writhe,
- negative (left-handed) twist \( \rightarrow \) right-handed ply with negative writhe.
The twist convention used here is that a positive kinematic twist rate, $\tau$, carries an outer longitudinal fibre of the rod into a right-handed helix. Although decreased in absolute magnitude by the ply formation, the handedness of the original twist is preserved. So, the positive twist of the left-hand ply is tending to overwind a right-handed DNA molecule: the DNA is then said to be positively supercoiled. Conversely, the negative twist of the right-hand ply is tending to underwind a right-handed DNA molecule: the DNA is then said to be negatively supercoiled. Note that these italicized descriptions are only used when discussing the more common right-handed DNA.

Note that a supercoiled DNA chain can, however, adopt either this interwound form or a toroidal form (a simple coil as exemplified by the NCP). The different rules that apply for this toroidal form are

- positive (right-handed) twist $\rightarrow$ right-handed coils with positive writhe,
- negative (left-handed) twist $\rightarrow$ left-handed coils with negative writhe.

To sharpen the topological concepts of twist, link and writhe, we imagine gluing the two ends of an initially straight rod together to form a closed loop. If we just bend the rod in a plane and glue the ends together without inserting any twist, it will adopt a circular shape, and all longitudinal fibres of the rod will be circles. Suppose, however, that having bent it in a plane and brought the ends together, we insert, at the last minute, a number of full turns of twist ($Tw$) just before gluing. We define this number as the link, $Lk$, positive if it induces positive $\tau$ in the planar ring. For as long as the glued ring remains planar, we have $Lk = Tw$.

Strictly, the link of a DNA plasmid (closed loop) will vary by integer increments because each sugar–phosphate chain of the double helix must join to itself: and in mathematical topology, the link is also normally taken to be integer. However, in the context of elastic rod theory (where the ends of a rod can be glued together at any angle), it is convenient to ignore this technicality and speak as if $Lk$ varies continuously.

Now $Lk$ is a topological invariant. If we get hold of the glued ring and distort it, in or out of the plane, in any way we choose, the link will not change. The total twist does, however, change, and the two are related by the following important result:

$$\text{link (} Lk \text{)} = \text{writhe (} Wr \text{)} + \text{twist (} Tw \text{)},$$

where the total twist, $Tw$, is the integral around the rod of the kinematic twist rate, $\tau$. The writhe, $Wr$, is best introduced as the number of (signed) crossings in a view, averaged over all views. The number of crossings in a single view is called the...
directional writhe, $Dr$. In the particular side view of a ply in figure 9, for example, a count of the number of crossings gives $Dr = 8$. In general, in calculating $Dr$, we must take the number of signed crossings, according to the standard right-hand rule of physics. Averaging $Dr$ over all views, we then obtain $Wr$. The use of these concepts in analysing the deformations of elastic rods can be seen in [121].

We finally say a few words about the use of an induced ply in magnetic tweezer experiments that are used to infer the stiffness characteristics of DNA molecules. Gluing one end of the molecule to a fixed surface, and the other end to a magnetic bead, a strong magnet can be manoeuvred to apply tension and link, as shown in figure 10.

Sensing the length $Z(n)$, a comparison with rod theory in figure 10b gives estimates of DNA stiffnesses. Placing the DNA in a solution containing topoisomerases (cutting enzymes), individual cuts made by the enzymes can be identified in the trace of $Z$ against time (figure 10c), e.g. Thompson [93]. This gives vital clues as to how the topoisomerase functions in its role of alleviating torsional stress.

References


Research. DNA information
A. A. Travers et al.


74 Schrödinger, E. 1944 What is life? The physical aspect of the living cell. Cambridge, UK: Cambridge University Press.


Review. DNA information


