



# A reappraisal of the form – function problem. Theory and phenomenology

Luciano Boi<sup>1</sup>

Received: 3 August 2021 / Accepted: 11 March 2022 / Published online: 26 April 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

This paper is aimed at demonstrating that some geometrical and topological transformations and operations serve not only as promoters of many specific genetic and cellular events in multicellular living organisms, but also as initiators of the organization and regulation of their functions. Thus, changes in the form and structure of macromolecular and cellular systems must be directly associated to their functions. There are specific classes of enzymes that manipulate the geometry and topology of complex DNA–protein structures, and thereby they perform many important cellular processes, including segregation of daughter chromosomes, gene regulation, and DNA repair. We argue that form has an organizing power, hence a causal action, in the sense that it enables to induce functional events during different biological processes, at the supra-molecular, cellular, and organismal levels of organization. Clearly, topological forms must be matched with specific kinetic and dynamical parameters to have a functional effectiveness in living systems. This effectiveness is remarkably apparent, to give an example, in the regulation of the genome functions and in cell activity. In more general terms, we try to show that the conformational plasticity of biological systems depends on different kinds of topological manipulations performed by specific families of enzymes. In doing so, they catalyze all those spatial and dynamical changes of biological structures that are suitable for the functions to be acted by the organism.

**Keywords** Form · Function · Geometry · Topology · Recombination · Epigenome · Cell activity · Global metabolism

## Introductory remarks

Among the general aims of this paper, we want to remind the followings. A first major motivation is the conviction that the integration of mathematics, physics, and molecular and cell biology will constitute the new frontier and challenge for twenty-first century science. The second relates to the fact that the exciting and appealing science in the twenty-first century is likely to evolve across, not within, traditional disciplines. Therefore, we will focus on the interfaces of mathematical methods and modeling with the physical and biological aspects of living systems. One of our main goals is to study the central role of multi-level and scale-change phenomena in the biological sciences (Jost 2019).

We aim at showing that several methods and techniques, especially from differential geometry and topology, are profoundly involved at different scales and various levels of organization in the physical and biological processes. We emphasize the need for developing new mathematical methods, models, and techniques suited to work out a topological–dynamical theory of the emergence of natural and living patterns and behaviors. For example, in our view, one particularly interesting task would consist of explaining to what extent the mathematical structure and spatial–temporal events that constitute the natural frame of living organisms may influence their bio-chemical, physiological, and metabolic organization and regulation.

In fact, there are effective mathematical models and techniques which can be used to describe several fundamental properties and behaviors observed in biological systems. Specifically, they may help to show that the complex topology and dynamics of DNA–protein complexes are closely linked to the multi-level epigenetic regulation and to the cell's spatial and functional organization. Finally, it will be important to emphasize that the geometrical structure and

---

✉ Luciano Boi  
lboi@ehess.fr

<sup>1</sup> École des Hautes Études en Sciences Sociales, Centre de Mathématiques (CAMS), 54, bd Raspail, 75006 Paris, France

topological form of nuclear components (DNA, nucleosome, chromatin, chromosome, etc.) play an important role in the cell differentiation (during an embryo's development) and organism growth.

Let us just mention an example which illustrate the connection between topological operations and biological processes. At the molecular- and supramolecular-level enzyme topoisomerases, which convert DNA from one topological form to another, appear to have a profound role in the central genetic events of replication, transcription, recombination, and repair (see Cozzarelli 1992; Wang 1996; Roca 1998). Moreover, certain topological mechanisms are involved in the fundamental biological process of the compaction (or condensation) of chromatin into the chromosome during the interphase and the metaphase.<sup>1</sup> (For more details, we refer to Hinde et al. (2012), and Dixon et al. (2016). Besides, it is time to suggest new mathematical methods relating to the cell's differentiation and their complex spatial organization during the different phases of the developments of the embryos.

<sup>1</sup> Recall that chromatin is achieved through the wrapping of DNA around a core of eight histone proteins at regular intervals along the entire length of the chromosome, forming the basic building blocks of the chromatin fiber, the nucleosomes (McGinty and Tan 2015). The nucleosomes are further compacted into high-order chromatin architecture, and organized into condensed compartments or heterochromatin domain and open compartments or euchromatin domain. Within the nucleus, histones provide the energy (mainly in the form of electrostatic interactions) to fold DNA. As a result, chromatin can be packaged into a much smaller volume than DNA alone. Chromosome compaction is on the order of several thousand-fold, yet these chromosomes have to be unraveled every cell cycle to be replicated accurately and the daughter chromosomes must be topologically unlinked to allow their separation and segregation into the daughter cells. During mitosis, although most of the chromatin is tightly compacted, there are small regions that are not as tightly compacted. These regions often correspond to promoter regions of genes that were active in that cell type prior to chromatin formation. During interphase (1), chromatin is in its least condensed state and appears loosely distributed throughout the nucleus. Chromatin condensation begins during prophase (2) and chromosomes become visible. Chromosomes remain condensed throughout the various stages of mitosis (2–5). Condensing chromatin is necessary not only for structural and functional (which we describe accurately in the main text), but also for physical reasons. There are proper physical properties that the condensation of chromatin into sturdy chromosomes must realize. Chromosomes must be stiff, robust, and elastic enough to withstand forces coming from pulling microtubules and cytoplasmic drags during mitosis to prevent damage and breaks caused by external tensions (Durickovic et al. 2013). Compaction status of chromatin is regulated by structural (spatial) and chemical modifications upon DNA sequences and histone proteins, such as DNA methylation (Suzuki and Bird 2008), histone acetylation, and methylation. Chromatin compaction regulates transcription activities, and impacts many genomic functions such as DNA replication, damage, and repair. Therefore, our capacity to explore chromatin architecture and its epigenomics states at molecular and macromolecular scales is essential to our understanding of functional significance of chromatin compaction status and elucidate many biological and anomalous processes.

From a more philosophical point of view, we think that it is essential to provide more global and dynamic mathematical ideas and models, and to rethink effectively the causal connection between topological form and biological function in living systems.

### Geometry of the DNA: the linking number and its connection with genomic processes

The principal goal of this paper is to highlight some key links between topology, physics, and biology, to show that topological operations (like knotting and surgery) and deformations (like embeddings and immersions) and dynamics take part in the living processes (Mazur 2004; Boi 2011a). I will limit myself to analyze some features of macromolecular structures like DNA–protein complexes, the chromatin, and the chromosome. All things I will speak about take place in the 3-dimensional space of living cells and particularly in the nucleus, which of course interact in many ways and at different levels with the whole cell, its cytoplasm, and the organelles.

This is of course a very partial view, an oversimplification, of what really happens in living organisms. Nevertheless, I guess that in many different contexts of biological sciences, we have to deal with the following problem<sup>2</sup>: how small local changes in a living system do affect the global behavior and response of the whole organism, and, conversely, to what extent the global metabolism of an organism can influence each of its specific functions? To answer this question, we need a clear picture of the most relevant spatial and temporal dimensions and spaces fitting biological phenomena.

Let us start by four observations and statements on the complexity of biological systems.

*Observation 1* We think that their study involves both the qualitative and quantitative employment and the simultaneous integration of different biological components, and of their relationships, as well. For example, the components may be proteins, while their relationships may be described by signal transduction pathways. The cellular processing is a complex-dynamic system with hundreds of thousands of bio-molecules interacting with one another to perform life's many functions. To fully understand the multi-layer information and organization “program” of life, a comprehensive description of protein–protein, cell–cell, cell–organism, and organism–environment interactions is required. Understanding how genes and their proteins products and cells and their

<sup>2</sup> This problem was addressed especially by Denis Nobel in the book *The Music of Life. Biology beyond the genome*, Oxford University Press, Oxford, 2006, and by Stuart Kauffman in its book *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, Oxford, 1993.

intra- and extra-interactions generate the complexity and diversity that we know as life is perhaps one of the greatest challenges of biological sciences (Scherrer and Jost 2007).

**Observation 2** The genome must be viewed as a complex structural system. In fact, recent theoretical studies and a huge amount of experimental data point toward the need for a profound change in our way of thinking about biological phenomena, and their modeling. Let us summarize some important findings:

1. In the last two decades, it has become more and more clear that the linear sequence map of human genome is an incomplete description of our genetic information and processing. This is because information on genome functions and gene regulation is also encoded in the way DNA string is folded up with proteins into chromosome within the nucleus. This allowed for the conclusion that *the biological information on living organisms cannot be portrayed in the DNA sequence alone*. In a post-genomic (epigenomic or proteomic?) era, the importance of chromatin–chromosome/epigenetic remodeling interface has become increasingly apparent.
2. The genome of eukaryotes is a highly complex system, which is regulated at (at least) five major (hierarchical or network-like?) levels: (a) the DNA molecule level, (b) the DNA–protein complexes and chromatin level (Bian and Belmont 2012; Boi 2009), (c) the regulation at RNA level, i.e., interactions of different RNAs or of RNAs with proteins, which make use of both geometric organization and a combinatorial code, and which is clearly among the most important regulatory steps, (d) the nuclear level, which includes the dynamics and three-dimensional spatial organization of the chromosome inside the nucleus, (e) the cell regulation in response to internal and external signals and factors, which is able to remodel the genome structure and function, and (f) the interactions between the global metabolism of organisms and their internal and external environments.
3. There is increasing evidence that such a higher order organization of chromatin structure contributes in an essential way to the regulation of gene expression and therefore to cell activity. Therefore, we must consider epigenetics to understand some features of our genome, its topological forms and the ways in which it functions (Waddington 1957; Villota-Salazar et al. 2016). The two properties are closely related. Epigenetics encompass the many processes that cannot be accounted for by the simple genetic code, and the term refers to extra layers of instructions, that is of biological organization and information (notably cellular, organismal, and environmental) that influences gene activity without altering the DNA molecule.

4. The information content of the DNA molecule is embodied in its sequence of paired nucleotide bases and it depends on how the molecule is twisted, tangled, or knotted. In other words, twisting, coiling, and knotting operations are able to enhance or to reduce the structural and physiological functions of the genome and cell nucleus. Moreover, it is now clear that the topological form of a DNA molecule, the structural modifications of the chromatin, and the spatial architecture of the chromosome influence the way in which DNA acts within the cell. These three levels of organization of the most fundamental nuclear components seem to be deeply related. Furthermore, their functions are controlled by the action of different complexes of regulatory factors and co-factors. Among these different families of protein regulatory complexes, the remodellers of chromatin structure play a fundamental role in replication and repair of DNA sequences and in the transcriptional activity of the entire genome (Ophl and Roberts 1978; Alberts 2003).

Before we go further, we need to introduce two mathematical concepts, which are central to our scope here. First, let us give the mathematical definition of the concept of *twist*, which plays a crucial role in almost all supramolecular and cellular processes. The topologist Max Dehn introduced a very far-reaching definition of the concept of twist (Dehn 1910). A *Dehn twist* is a certain type of an orientation-preserving homeomorphism of a surface. Suppose that  $c$  is a simple closed curve in a closed, orientable surface  $S$  of genus  $g$ , precisely of the surface obtained by cutting the surface along  $c$  (a circle), rotating, and gluing back (Seifert 1935). (For example, one may imagine a circle representing one of  $2g$  generators of  $H^1(S, \mathbb{Z})$ .) Let  $A$  be a tubular neighborhood of  $c$ . Then,  $A$  is an annulus, homeomorphic to the Cartesian product of a circle and a unit interval  $I$ :  $c \subset A \cong S^1 \times I$ . Give coordinates  $(z, t)$  where  $z$  is a complex number of the form  $e^{i\theta}$  with  $\theta \in [0, 2\pi]$ , and  $t \in [0, 1]$ . Let  $f$  be the map from  $S$  to itself which is the identity, outside of  $A$  and inside of  $A$  we have  $f(z, t) = (ze^{2\pi it}, (z, t))$ . Then,  $f$  is a *Dehn twist* about the curve  $c$ . Dehn twist can also be defined on a non-orientable surface  $S$ , provided that one starts with a 2-sided simple closed curve  $c$  on  $S$ . Dehn twists appear in a number of basic constructions in low-dimensional topology. This mainly stems from the so-called “Dehn-Lickorish theorem”, stating that Dehn twists give rise to generators for the mapping class group of compact oriented surfaces (Birman 1974). The precise statement is.

**Theorem** (Dehn-Lickorish). *The Dehn twists generate the mapping class group of  $S$ , of orientation-preserving homeomorphisms considered modulo isotopy. (In fact, Lickorish described  $3g-1$  explicit embedded circles for a surface  $S$  of genus  $g$  whose corresponding twists give the generators.)*

An important conceptual issue here is that a Dehn twist does not change the topology of the surface itself, but only how the generators of its first homology are represented. For instance, the presentation of closed orientable 3-manifolds in terms of framed links in the 3-sphere relies crucially on this fact (Zeeman 1960). For another instance, Dehn twists appear as monodromy around critical points of Lefschetz fibration and thus provide a combinatorial approach to the study this interesting class of 4-manifolds. (For a detailed presentation of this subject, we refer to Rolfsen 1976; Lickorish 1997; Burde and Zieschang 2003).

The second concept is that of *linking number* (Kauffman 2001; Sergei 2001; Spera 2006). In mathematical terms, the linking of two closed curves is a topological property: no matter how the curves are deformed (pulled, twisted, and so on), as long as neither one is broken, they will remain linked in exactly the same way. The linking number, here denoted by  $Lk$ , is defined as a signed integer that describes a property of two closed curves in space. To separate a pair of curves without actually cutting them, the value of  $Lk$  must be 0 (although the converse is not always true). If the curves in question are the edges of a closed ribbon with  $n$  turns in it, their linking number will remain unchanged when the ribbon is deformed (Fuller 1978). The linking number of two smooth, regular and oriented curves in space is one of the basic invariants, which gives topological information<sup>3</sup> about them; it tells us how many times one curve winds around another. These curves can be deformed by some kind of repositions called *Reidemeister moves* (see section “Some topological concepts” for more mathematical details). Suppose that the intertwined curves  $\gamma_1$  and  $\gamma_2$  are represented by an oriented 2-component link diagram  $L$ , attach a sign (+1 or -1) to each crossing. Then, the linking number,  $Lk(\gamma_1, \gamma_2)$ , is the sum of these signs over all crossings of  $\gamma_1$  with  $\gamma_2$ . It can be shown that the linking number is invariant under Reidemeister moves (Reidemeister 1932; Boi 2006). That is, if we take a given diagram  $D$  of the curves  $\gamma_1$  and

$\gamma_2$  and change it to a new diagram  $D'$  by applying one of the Reidemeister moves, then the linking number calculation for  $D$  will be the same as the calculation for  $D'$ . The calculation is unaffected by the first Reidemeister move, because self-crossing of a single curve does not figure in the calculation of the linking number. The second Reidemeister move either creates or removes two crossings of opposite sign, and the third move rearranges a configuration of crossings without changing their sign.

These facts are the first step in the effective application of algebraic and geometric topology to the study of knots and links (Kauffman 1987; Adams 2000; Boi 2006). The concept of linking number and its successive findings has a long and interesting history (originated in the Gauss's studies on the magnetic potential and the topological investigations made by Listing followed by the successive developments by Thomson, Maxwell and Tait in the second half of the XIX century) and there are a number of ways to define it, many considerably more complicated than the sum of diagrammatic signs. Some of these different though equivalent definitions are discussed in Kauffman (2005) and Ricca and Nipoti (2011). There are at least three many interpretations of the linking number, namely, in terms of *degree*, *signed crossings* and *intersection number*. To give a straightforward mathematical definition of the linking number, consider an oriented diagram  $D_\nu(L)$  of the (tame) link  $L = \gamma_1 \sqcup \gamma_2$ , obtained by projecting  $L$  along  $\nu$  onto the plane, allowing under- and over-crossings. Let  $D_\nu(L)$  be a god projection of  $L$ , that is one for which the standard projection has nodal points of multiplicity at most two. We assign to each apparent crossing  $c$  of  $\gamma_1 \cap \gamma_2$  the number  $\varepsilon(c) = \pm 1$  according to the standard convention. We have the following definition.

**Definition 1** The *linking number* of  $Lk(\gamma_1, \gamma_2)$  of  $\gamma_1$  and  $\gamma_2$  is defined by

$$Lk(\gamma_1, \gamma_2) : = \frac{1}{2} \sum_{c \in \gamma_1 \cap \gamma_2} \varepsilon(c).$$

This number has some very striking properties, the most important of which is that the linking number  $Lk(\gamma_1, \gamma_2)$  is an invariant of  $L$ , that is, it is the same for two or more diagrams of  $L$ .

This mathematical result, namely the fact that linking number is a numerical invariant that describes how many times two closed curves are entangled in three-dimensional space, find a natural application to biology, since the linking number is a topological property of DNA string. Precisely, it is a sum of twists and writhes. In short, the *twist* is the number of times a DNA-strand turn around the other strand. And the *writhe* is the number of times DNA double helix is crossed, coiled over each other or the number of times one strand wrap around the other strand. We can also say that the twist is the number of helical turns in the DNA string, and

<sup>3</sup> *Topological information* is information about a knot or link that does not depend upon the material from which it is made and is not changed by stretching or bending that material so long as it is not torn in the process. We do not want the knot to break up when the material undergoes some change in one or more of its physical parameters or to disappear in the course of such a stretching process by slipping over one of the ends of the rope. Precisely, topological information is invariant by deformation. Topological information about knots and links can be obtained from different sources. (1) From their diagrammatic representation and the associated Reidemeister moves. (2) From their numerical, algebraic, and topological invariants, starting with the most basic like the linking number to other more complete and powerful invariants like the Jones polynomial. (3) From quantum groups and quantum invariants of 3- and 4-manifolds. (4) From statistical mechanic models and critical phenomena. (5) From macroscopic physics, especially fluid mechanics and hydrodynamics. (6) From molecular biology, particularly from the replication and recombination processes.

the writhe is the number of times the double helix crosses over itself (these are supercoils) (White et al. 1988). Extra helical twists are positive and lead to positive supercoiling, while subtractive twisting causes negative supercoils (see sections “DNA–histone complexes and the packaging of chromatin” and “Topological enzymology: linking number, supercoiling, and topoisomerases” for a comprehensive discussion of this topic).

### Some remarks about the organization of the chromosome

In the nucleus, individual chromosomes occupy discrete topological territories. Examining the spatial organization (evolving in time) of human chromosomes and genes in the nucleus appears to be very important. It seems that this organization is changed, for example, during development and in certain diseases. Consequently, the way the human chromosome is topologically organized might influence how abnormal chromosomes are formed. Using whole chromosome painting probes and fluorescence in situ hybridization (FISH), a territorial organization of interphase chromosome has been demonstrated (Cremer et al. 2004). Chromosome territories have irregular shapes and occupy nuclear positions with little overlap. In general, gene-rich domains of chromosome are located in the nuclear interior, while gene-poor chromosome domains are more situated in the nuclear periphery. In agreement with this, non-transcribed sequences were predominantly found at the nuclear periphery, with active gene regions tended to localize on chromosome surfaces exposed to the nuclear interior or on loops extending from the territories (see Cremer et al. 2004; Misteli 2007). Chromosomes have essentially two structurally and functionally distinct territories: euchromatin and heterochromatin. Heterochromatin, which is mostly accumulated adjacent to the nucleus envelope, is highly condensed, gene-poor, and transcriptionally silent, whereas euchromatin, which is rather dispersed in the whole interior of the nucleus, is weakly condensed, gene-rich, and much more transcribed. The two-form topological organization of chromatin is functionally important also, because heterochromatin maintains the structural integrity of the genome and allows the regulation of gene expression (Ochs et al. 2019), while euchromatin allows the genes to be transcribed and variation to occur within them.

These experimental findings support the concept of a *functional nuclear space*, the inter-chromosomal domain compartment (ICD). According to ICD model, the interface between chromosome territories is more easily accessible to large nuclear complexes than to regions within the territory. More recently, it has been proposed that chromosome territories are further organized into 1-Mb domains, extending

the more accessible space open intra-chromosomal regions surrounded by denser chromatin domains. Using high-resolution light microscopy, an apparent bead-like structure of chromatin can be visualized in which around 1-Mb domains of chromatin are more densely packed into an approximately spherical sub-compartment structure with dimensions of 3000–4000 nm. (See Cremer et al. 2004; Ramam et al. 2016).

### DNA–histone complexes and the packaging of chromatin

The key distinguishing characteristic of the eukaryotic genome is its tight packaging into chromatin, a hierarchically organized complex of DNA and histone and non-histone proteins. How the genome operates in the chromatin context is a central question in the molecular genetics of eukaryotes. The chromatin packaging consists of different levels of organization. Every level of chromatin organization, from nucleosome to higher order structure up to its intranuclear localization, can contribute to the regulation of gene expression, as well as affect other functions of the genome, such as replication and repair. Concerning gene expression, chromatin is important not only because of the accessibility problem it poses for the transcription apparatus, but also due to the phenomenon of chromatin memory, that is, the apparent ability of alternative chromatin states to be maintained through many cell divisions. This phenomenon is believed to be involved in the mechanism of epigenetic inheritance, an important concept of developmental biology.

Today, we know that DNA is topologically polymorphic (Strick et al. 1998; Zhurkin and Norouzi 2021). The overwound or underwound double-helix can assume exotic forms known as *plectonemes*, like the braided structures of a tangled telephone cord, or *solenoids*, similar to the winding of a magnetic coil.

1. Plectonemically supercoiled DNA is unrestrained and frequently branched, while toroidal supercoils is restrained by proteins and it is more compact (Boles et al. 1990). The extended thin form of plectonemically supercoiled DNA offers little compaction for cellular packaging, but promotes interaction between cis-acting sequence elements that may be distant in primary structure.
2. DNA can be either positive or negatively supercoiled. In particular, eukaryotic DNA is negatively supercoiled in and around genes, and it is transiently negatively supercoiled behind RNA polymerase during transcription.
3. Negative supercoiling favors DNA–histone association and the formation of nucleosomes, the first step in packaging DNA. Because the solenoidal DNA wrapping

around a nucleosome core creates about two negative supercoils, it is understandable that the DNA that fulfills this topological prerequisite will more easily form nucleosome.

4. These tertiary structures have an important effect on the molecule's secondary structure and eventually its functions. For example, supercoiling induces destabilization of certain DNA sequences and allows the extrusion of cruciform or even the transcriptional activation of eukaryotic promoters. Another essential process, DNA transcription, can both generate and be regulated by supercoiling (Muskhelishvili and Travers 2016).

During replication, the chromosome needs to be partitioned and the two strands of DNA must be continuously unlinked. The topoisomerases that accomplish this might instead be expected to entangle and knot chromosomes because of the huge DNA concentration in vivo. There are actually several factors that solve this problem and contribute to the orderly unlinking of DNA. A major contributor to chromosome partitioning is the condensation of daughter DNA upon itself soon after replication. DNA condensation is due primarily to supercoiling. Another factor promoting chromosome partitioning is that the type II topoisomerases of all organisms do not just speed up the approach to topological equilibrium, but actually change the equilibrium position. They actively remove all DNA entanglements. This requires that topoisomerases sense the global conformation of DNA, even though they interact with DNA only locally.

In fact, topoisomerases achieve this, because, by positioning themselves at sharp bends in DNA, they carry out net disentanglement of DNA. They act, in a way, like topological operators with a functional target. An equal partner to the topoisomerases in chromosome segregation is the helicases. They seem to convert the energy of ATP hydrolysis into unwinding DNA. All the enzymes that play critical roles in DNA unlinking and chromosome segregation, topoisomerases, helicases, and condensins, are motor proteins. They use the energy of ATP hydrolysis to move large pieces of DNA over long distances.

The previous discussion can be summed up by saying that supercoiling accomplishes three essential functions (Brunello et al. 2012).

1. First, (–) supercoiling promotes the unwinding of DNA and thereby the myriad processes that depend on helix opening.
2. The second essential function of supercoiling is in DNA replication. For replication to be completed, the linking number of the DNA,  $Lk$ , must be reduced from its vast (+) value to exactly zero. In bacteria, DNA gyrase

introduces (–) supercoils and thereby removes parental  $Lk$ . DNA gyrase is unique among all topoisomerases and it is the only enzyme that is able to negatively supercoil the double helix.

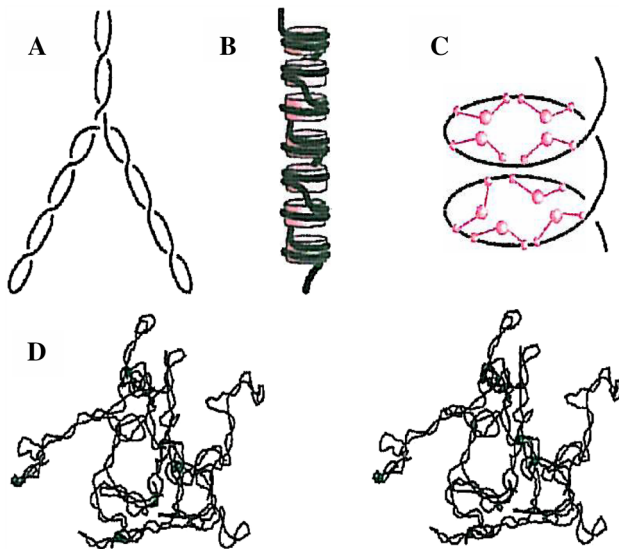
3. The third essential function of supercoiling is conformational. DNA manifests the difference between the relaxed and naturally occurring values of  $Lk$  by winding up into supercoils. These supercoils condense DNA and promote the disentanglement of topological domains. This can be accomplished equally well by (–) or (+) supercoiling.

Let us still underline two important facts. First, the promotion of decatenation by supercoiling has also been directly demonstrated in vivo. Second, the volume occupied by a supercoiled molecule is much smaller than that of a relaxed DNA. This difference in volume is due mostly to the formation of superhelical branches. Indeed, supercoiled DNA branches and bends itself into a ball. The decrease in chromosomal volume by supercoiling reduce the probability that the septum will pass through the chromosome during cell division.

It seems clear that supercoiling plays a fundamental role in the condensation of the double helix and that this condensation is responsible for DNA unlinking and chromosome partitioning. Supercoiling results from topological strain and the contortion of DNA by proteins, notably the nucleosomal histone octet and the structural maintenance of chromosomes (SMC) proteins. There are three ways, actually experimentally observed in vivo, in which condensation of the DNA–protein complexes into chromatin by supercoiling occurs, and to each of them corresponds a topological model for explaining the compaction of chromosomes in the cell's nucleus.

Let us now describe in detail the three ways through which supercoiling is performed (Fig. 1).

1. (–) Supercoiling by gyrase compacts the chromosomes such that random passages by topoisomerase IV disentangle them. In particular, topoisomerase IV is responsible for decatenation of DNA.
2. With the second type of condensation via supercoiling, that is by folding around the core histones proteins (i.e., the nucleosome), DNA is compacted in independent successive stages, such that the total compaction is the product of compaction in each stage. The first stage of this compaction is via solenoidal wrapping of DNA in the nucleosome. Although the compaction achieved is modest, the nucleosome provides a fundamental structure for genome organization and function. The structure of a nucleosome reveals a scaffolding that forces the DNA to adopt ordered solenoidal supercoils.



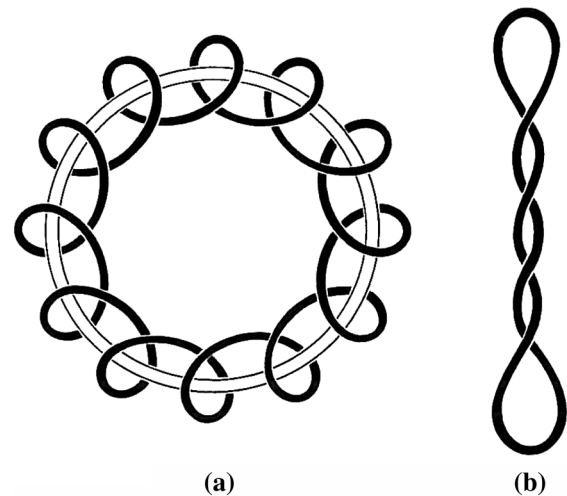
**Fig. 1** Comparison of three in vivo types of DNA compaction by supercoiling. **a** Free (–) supercoils twist DNA into a right-handed plectonemic superhelix. **b** Wrapping around the histone octamer compacts DNA by forming left-handed solenoidal supercoils. **c** SMC proteins, such as *Xenopus* 135 condensin (schematized as red ball and stalk structures), effect global DNA writhe by forming large (+) solenoidal supercoils. **d** This is a stereo image of a 25-kilobase (kb) (–) supercoiled DNA generated by a Metropolis Monte Carlo simulation. **a** and **c** represent approximately 2 kb of DNA (700 nm) at 200,000-fold magnification, whereas **b** is only 1.5 kb of DNA (500 nm) but at four-fold greater magnification. **d** is at 100,000-fold magnification

3. The third type of compaction *cum* supercoiling, that by condensin,<sup>4</sup> is needed for the formation of mitotic chromosomes from the open interphase forms (Hirano 2016).

Supercoil can have an interwound or a toroidal 3-D shape.

(1) The circular DNA (that is, with the ends of the molecule fixed) consists of a series of open spirals that wind around an imaginary ring or toroid; this kind of supercoiling is known as *toroidal*. However, the circular can also wind above and below itself several times, and this kind of supercoiling is called *interwound*. (Vologodskii 1992). In

<sup>4</sup> *Condensins* are large protein complexes that play a central role in chromosome assembly and segregation during mitosis and meiosis in the three domains of life. They display highly characteristics, rod-shaped structures with SMC (structural maintenance of chromosomes) ATPases as their core subunits and organize large-scale chromosome structure through active mechanisms. Most eukaryotic species have two distinct condensins' complexes whose balanced usage is adapted flexibly to different organisms and cell types. One has observed both conserved features and rich variations of condensin-based chromosome organization. *Cohesins* are another representative class of eukaryotic SMC protein complexes. They play a central role in sister chromatid cohesion during mitosis and meiosis. Recent studies highlight their participation in gene regulation, in close collaboration with the insulator CTCF.



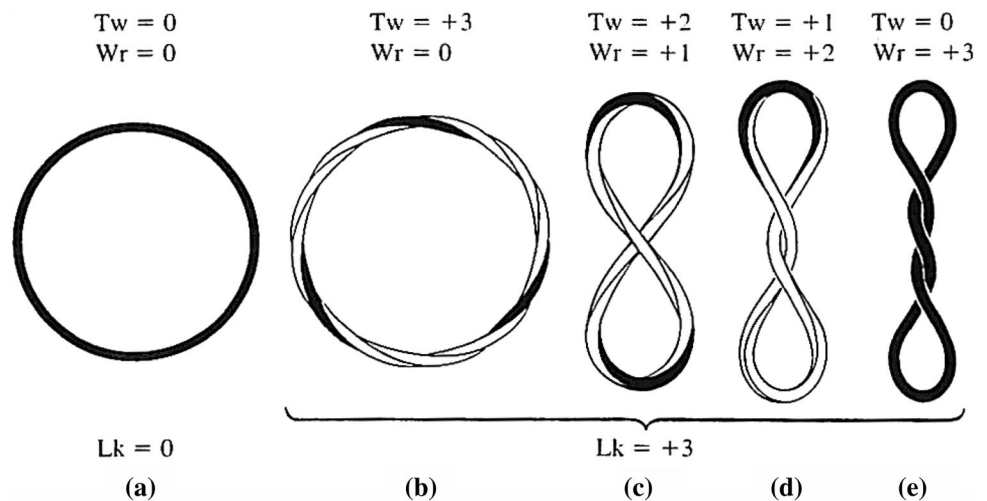
**Fig. 2** Two general varieties of DNA supercoil. In **a**, the DNA coils into a series of spirals about an imaginary toroid or ring (shown here by open lines); and so, this kind of wrapping is known as “toroidal”. In **b**, the DNA crosses over and under itself repeatedly; and therefore, this kind of wrapping is known as “interwound”

practice, real DNA supercoils may contain portions of both the toroidal and interwound geometries. Thus, where certain parts of the DNA are highly curved, on account of either the base sequence or due to wrapping around a protein, one may find toroidal structures, since the DNA in a toroidal supercoil is highly curved throughout. Alternatively, if such curved portions of the DNA are not very long, they may locate themselves at the two strongly curved end-loops of an interwound supercoil, as shown on the left and the right in Fig. 2. Sometimes, the interwound and toroidal geometries may occur together, as in the looped-linear DNA. Linear DNA molecule into loops generates end-restraint at the base of every loop, if the two ends are attached to some support of “scaffold”. This kind of looped-linear arrangement is thought to be typical of the chromosomal DNA found in higher organisms. On a small scale, within any loop, the coiling is toroidal on account of the wrapping of DNA around protein spools; but on a large scale, over the full length of any loop, the structure is interwound. You often see this kind of arrangement in “hold-time” telephone cords, if people habitually rotate the handset.

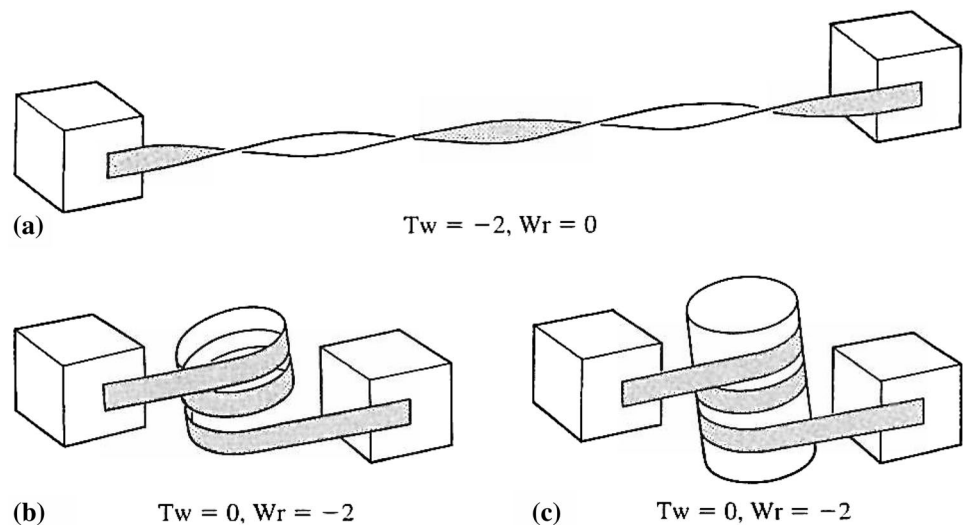
In general, supercoiled DNA has the shapes seen in Fig. 2, because it either has more turns of twist, or fewer turns of twist, than the underlying, relaxed, right-handed double helix from which it is made. DNA with more than the natural number of turns is known as *overwound*, while DNA with fewer than the natural number of turns is known as *underwound*.

Now, what are the relative stabilities of these two forms of DNA supercoiling? In other words, when (that is, in which bio-chemical and physical conditions) will a DNA molecule

**Fig. 3** Five closely related circular DNA molecules: **a** and **b** show open circles, while **c–e** show interwound supercoils. The DNA in its stress-free, relaxed form is drawn as a rubber rod of square cross-section, with one face black



**Fig. 4** A highly twisted ribbon will collapse spontaneously into part of a toroidal supercoil. In **a**, the two ends of the ribbon are held apart by their attachment to blocks, so that  $Tw = -2$ . In **b**, the blocks move together, so that the ribbon can collapse to  $Wr = -2$ . In **c**, a cork or protein spool stabilizes the shape of the ribbon shown in **b**



be interwound, and when will it be toroidal? The interwound shape is usually very stable, and most underwound or overwound DNA molecules will naturally adopt an interwound shape, in the absence of other forces. However, the proteins that associate with DNA in living cells can sometimes change the situation dramatically, and favor the toroidal over the interwound form by wrapping the DNA around themselves (see below for further details). Note, however, that the preferred interwound structure of DNA molecules in cells is somewhat similar to the idealized shape in Fig. 3e (but with a linking number  $Lk$  of the opposite sense, which means that these DNA molecules are underwound, with  $Lk$  negative), since  $Wr = 0.9 Lk$ , and  $Tw = 0.1 Lk$ . In other words, the DNA which has been underwound finds it more favorable energetically to cross over itself repeatedly, than to alter its twist.

Let us describe the hypothetical following model of supercoiling. Consider for example the cork which has

been inserted between the two turns of the ribbon shown in Fig. 4c. This cork represents a typical protein “spool” around which the DNA can wrap, and around which it does wrap in a left-handed sense in the chromosomes of higher organisms. If the DNA or ribbon in Fig. 4c were to be cut free from the two blocks at either end, it would stay wrapped around the “sticky” protein spool; whereas if it were cut free in the absence of a spool, as in Fig. 4b, it would immediately spring back into a straight configuration.

When we isolate DNA in the laboratory in pure form from any kind of cell or cells, at some point in the procedure, we must strip off the proteins around which the DNA was originally wrapped, without breaking either of its two double-helical strands. In other words, we must remove the cork from the arrangement shown in Fig. 4c, without cutting the DNA free from either of its two end-blocks. Naturally, the ‘naked’ DNA will first spring out to the highly twisted form shown in Fig. 4a, and then, it can collapse into



an interwound supercoil, because it has lost the curvature which stabilized the toroidal form.

Therefore, we can expect to see highly interwound supercoils in the preparations of pure DNA which we make from living cells, after removal of various proteins. Incidentally, this is why, DNA supercoils in Nature are usually underwound rather than overwound: the DNA always coils around proteins in the cell nucleus in the form of a left-handed toroidal spiral, giving negative  $Lk$ . In the next section, we will be especially concerned with some important topological and biological properties of supercoiling.

## Modeling the folding of chromatin

Among the different hypothetical models that have been proposed over the last years for the folding of the chromatin fiber during interphase, the so-called *radial-loop model* seems to us the most suitable for explaining the formation of the 30-nm solenoid structure. We suggested, specifically, a theoretical model by applying some methods and techniques from geometric topology and algebraic geometry.

The geometrical model we suggested might fit well with the 3-dimensional packing process of chromatin, first, into a 30-nm extended scaffold-associated form. In fact, the condensation of metaphase chromosome results from several orders of folding and coiling of the 30-nm chromatin fiber. For example, electron micrographs of histone-depleted metaphase chromosome from HeLa cells reveal long loops of DNA anchored to a chromosome scaffold composed of non-histone proteins. This scaffold has the shape of the metaphase chromosome and persists even when the DNA is digested by nucleases. Mega-base long loops of the 30-nm chromatin fiber are thought to associate with flexible chromosome scaffold, yielding an extended coiling of the scaffold into a helix, and further packing of this local structure produces the highly condensed structure characteristic of metaphase chromosome.

The topological complexity of DNA is strongly related to its biological meaning (White 1989). Let us first emphasize an important point, namely, that the complex topology of DNA is essential for the life of all organisms (Buck 2009). In particular, it is needed for the process known as DNA replication, whereby a replica of the DNA is made and one copy is passed on to each daughter cell. The most direct evidence for the vital role played by DNA topology is provided by the results of attempts to change the topology of DNA inside cells. Two related questions arise immediately from the recognition that DNA topology is essential for life. How did the complex topology of DNA evolve, and why is it so important for cells? DNA is the only molecule in cells that has a complex topology.

Type I topoisomerases of the DNA molecule, which cut one strand at a time, can carry out several topological

operations (Forterre et al. 2007).<sup>5</sup> By cutting one strand of a supercoiled DNA ring, the type I enzyme can put the ring into the relaxed state. It can tie a single-strand ring into a knot. The knot is tied when the simple-strand ring crosses over itself. If the two loops formed in this way are pulled together, the enzyme can cut one loop and pass the other loop through the opening. When the break is sealed, the ring is sealed in a knot. The type I enzyme can also interlock two single-strand rings. If the rings have complementary base sequences, a double-helix results. Although the operations seem quite different, each requires that a strand be broken, a segment of DNA be passed through the break and the break be resealed.

The evolution of proteins has taken a different course. Proteins also naturally subdivide into domains and thus local knots or links could readily occur, but they do rarely, although different types of pseudoknots have been recently observed in proteins patterns. Besides, no proper knots, catenanes, or supercoiling have been found so far in RNA, polysaccharides, or lipids. However, in view of recent works by C. M. Reidys and his coworkers (see Huang and Reidys 2015, 2016), it must be said that RNA may present pseudoknot structures; pseudoknots can be defined as a bipartite helical structure formed by base pairing of the apical loop in the stem-loop structure with an outside sequence. RNA pseudoknots are structural motifs in RNA that are increasingly recognized in viral and cellular RNAs (Theimer et al. 2005). More precisely, morphologically they are double-stranded helices that participate in the formation of different folding topologies and constitute the major fraction of RNA structures. Pseudoknots are formed upon base pairing of a single-stranded region of DNA in the loop of a hairpin to a stretch of complementary nucleotides elsewhere in the RNA chain. Reidys and Huang studied specific topological properties of RNA structures, particularly RNA contact structures with cross-serial interactions that are filtered by their topological genus, and then, they revealed that RNA secondary structures are topological structures having genus zero. The authors of these studies showed that a topological RNA structure can be obtained by fattening the edges of a contact structure into ribbons. The shape of a topological RNA structure is found by collapsing the stacks of the structure into single arcs and by removing any arcs of length one, as well as isolated vortices. Accordingly, a shape contains the key topological information of the molecular conformation, and the authors demonstrated that for fixed topological genus, there exist only finitely many such shapes. Furthermore, it must be stressed that pseudoknots constitute integral parts of the RNA structure essential for various cellular activities. Among many functions of pseudoknotted RNAs is feedback regulation of

<sup>5</sup> Type IB topoisomerases can facilitate DNA rotation in either direction, and they can relax negative or positive supercoils.

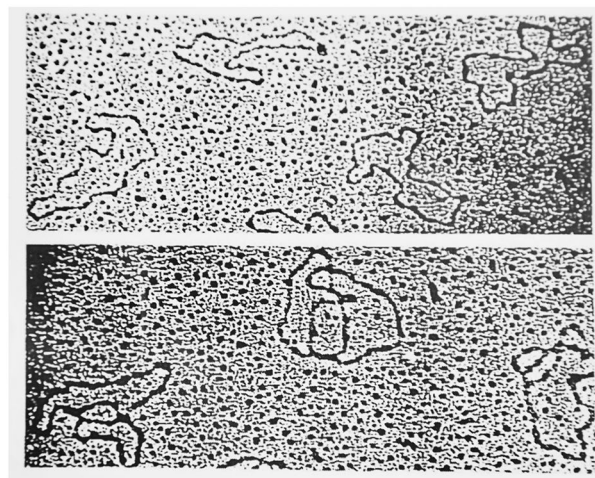
gene expression, carried out through specific recognition of various molecules (see Peselis and Serganov 2014).

The protein folding is one of the most important problems of the biological sciences (Gromov 2011; Flapan et al. 2019). It presents a high degree of structural complexity and of functional complexity, as well. Results from several recent studies in molecular and cellular biology and in mathematical biology clearly show that these two kinds of complexity are deeply related and that, in some sense, they act cooperatively to assure an efficient regulation of the genome and the epigenome and to preserve a certain stability of biological structures and functions (Carbone and Gromov 2001; Kitano 2004). Let us now remark that the standard approach in the study of proteins consisted in the study of one protein at a time. However, this approach showed its limits, and, therefore, we want point to two important hints of research: (1) biological function appears to be more a correlate of macro-molecular geometry than of chemical detail. (2) Any effective picture of protein structure must provide at the same time a model for the common character of all proteins as exemplified by their many chemical and physical similarities, and for the highly specific nature of each protein type.

The protein folding problem can be summarized in the following questions: How does a protein's amino acid sequence dictates its 3-D structure?<sup>6</sup> (1) The folding code: For a given sequence, what balance of interaction forces dictate the structure? (2) The folding process: What routes/pathways are used to reach the native structure quickly? (3) Protein structure prediction: computational predict native structure and folding pathway from a given sequence?

It is important to stress the topological determinants of protein folding. Indeed, for some protein, one can show that topological properties of protein conformations determine their kinetic ability to fold. One speaks of a macroscopic measure of the protein contact network topology, the average graph connectivity, by constructing graphs that are based on the geometry of protein conformations. It has been found that the average connectivity is higher for conformations

<sup>6</sup> Recall that in a protein, individual amino acids constituting the primary sequence interact with one another to form secondary structures such as helices and like-sheets surfaces. Next, individual amino acids from distant parts of the primary sequence can intermingle via charge-charge, hydrophobic, disulfide, or other interactions, and the formation of these bonds and interactions will serve to change the shape of the overall protein; this typical and complex folded structure corresponds to its tertiary structure. In other words, tertiary structure is the three-dimensional structure of a protein. Precisely, the tertiary structure of proteins deals with how the local structures are put together and ordered in space following certain geometric and combinatorial rules and codes. For example, the  $\alpha$ -helices may be oriented parallel to each other or at right-angles. Therefore, the tertiary structure refers to the folding of the different segments of helices, sheets, turns, and the remainder of the protein into the native three-dimensional structure.



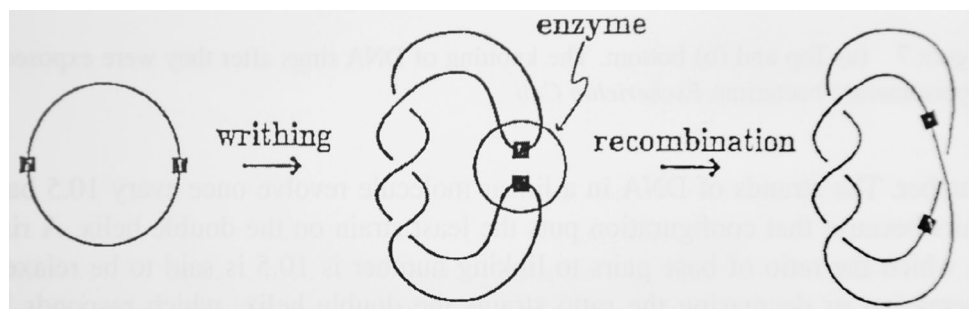
**Fig. 5** Top and bottom: the knotting of DNA rings after they were exposed to a topoisomerase bacterium *Escherichia coli*. Type I topoisomerase relax DNA by nicking and then closing one strand of duplex DNA. Type II topoisomerases change the DNA topology by breaking and rejoining double-stranded DNA

with a high folding probability than for those with a high probability to unfold. As a protein unfolds, it encounters dynamic constraints that emerge as a consequence of its being folded into a particular low-resolution structure or topology. For example, it often occurs that parts of a protein are entangled or wrapped within its interior, and for these “frustrated” parts to unfold requires the rest of protein to reorganize and at least partially unfold first. At this level of resolution, topological constraints can impose a time order on unfolding events, and occasionally, this order can be recognized in a protein's actual nucleation process or folding “pathway” despite the extreme complexity of its interactions.

### Topological enzymology: linking number, supercoiling, and topoisomerases

Before we go further, it is now important to describe some facts about topoisomerases. Their properties and action define what it can be called *topological enzymology*. They are enzymes which change the linking number of DNA strands; therefore, they have an important role in the central genetic events of DNA replication, transcription, and recombination. The DNA in the cell knots and unknots ties and unties itself according to a definite scheme. Knots and links appear during replication and recombination. Certain topoisomerases, which behave like topological entities in living organisms, are responsible for the knotting and unknotting. More precisely, they are able to cut a strand of DNA at a particular point, grasp another strand, pass it through the opening, and then close the opening. In other words,

**Fig. 6** The writhing process of the DNA molecule. The action of the enzyme topoisomerases type II consists in cutting both ends of the link and then recombining them by joining the two strands



these enzymes replace over-crossing by under-crossing. The tying of knots in rings of DNA is one of the capabilities of these enzymes. The ring can assume a number of topological configurations. The conversion of the DNA ring from one configuration to another is catalyzed by topoisomerases.

**Example** (see Fig. 5). Consider a single-strand DNA rings from a virus known as bacteriophage, which infects bacteria. What one observes of the rings, after they were exposed to a topoisomerase from the bacterium *Escherichia coli*, is then that, by cutting the DNA strand, passing a segment of the rings through the break and rejoining the cut ends, the enzymes has tied a knot in each ring. In fact, the process of breaking, passage and resealing is essential to the action of all topoisomerases. Some of the enzymes, type-I, cut a single strand of DNA; others, type-II, cut both strands of a double helix.

DNA is not at all a linear molecule and it exists in different spatial and functional states. In fact, it goes through different kinds of modifications of its shape during a cell cycle—that is the series of events that take place in a cell as it grows (interphase) and divides (mitosis), and these changes affect its functions.

Supercoiling is a fundamental geometrical state of the DNA duplex whose variations induce significant changes in the physiology of the molecule. It is also a process which displays a complex dynamic relating the plastic deformability of the molecule to its functional changes. Supercoiling of a double-strand DNA ring deforms the ring into a more twisted and compact shape. The shape of a DNA ring is strongly affected by the number of time one strand goes around the other, that is by the *linking number*. Since it is a topological quantity, it cannot be altered, while the strands are intact regardless of how the ring is pulled or twisted. If the strands are cut, however, and then rotated in the direction opposite to that of the twist of the helix, the helix unwinds. When the cut ends are rejoined, the number of rotations that have been made decreases the linking number. The strands of DNA in a linear molecule revolve every 10.5 base pairs because that configuration puts the east strain on the double helix. A DNA ring in which the ratio of base pairs to linking number is 10.5

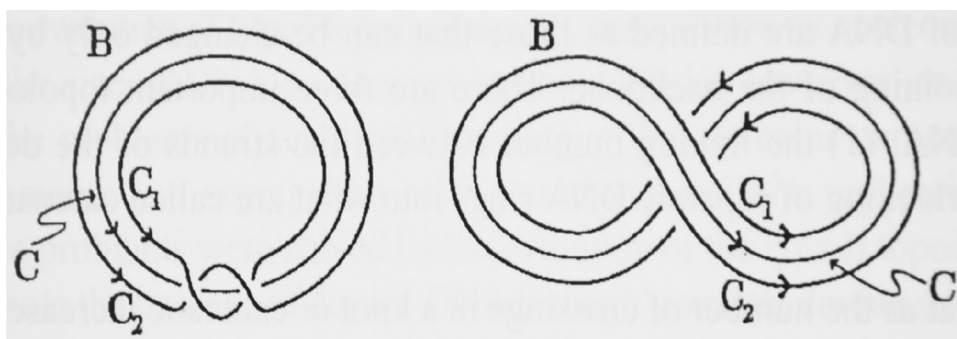
is said to be *relaxed* (that is, non-supercoiled). Increasing or decreasing the ratio strains the double-helix, which responds by supercoiling. In other words, a DNA molecule is sensitive to the variations of its topology. Reducing the linking number causes negative super coiling; raising the linking number leads to positive supercoiling.

Thanks to its topological properties, DNA is malleable and deformable. This property might distinguish living soft matter from non-living solid matter. This flexibility and topological deformability influence the biological functions of a double-helix. In fact, the molecule can move about in the space of the cell's nucleus and transform itself into several shapes without losing structural stability and energetic optimal state. This movement is two-fold: the three-dimensional two-stranded helical structure of DNA molecule can extend and compact. (1) The extended (unfolded) conformation DNA is especially required for replication. (2) DNA compaction inside cells occurs by successive order of coiling. A DNA double-helix is compacted in about four successive steps. The first step is the formation of the chromosome. The nucleosomes are coiled to give the final form, called a chromosome. In the phases of this processes, i.e., recombination, the knot type of DNA is changed (Figs. 6, 7).

Among the proteins involved in DNA replication are several that change the topology of DNA (see Lodish et al. 2000): helicases, which can unwind the DNA duplex, thereby inducing formation of supercoils, and topoisomerases, which catalyze addition or removal of supercoils. Type I topoisomerases relax DNA (i.e., remove supercoils) by nicking and closing one strand of duplex DNA. Type II topoisomerases change DNA topology by breaking and rejoining double-stranded DNA. These enzymes can introduce or remove supercoils and can separate two DNA duplexes that are intertwined. Topoisomerases are important both in growing fork or replication fork<sup>7</sup> movement and

<sup>7</sup> By the terms of *replication fork*, one designs a site in double-stranded DNA at which the template strands are separated and addition of deoxyribonucleotides to each newly formed chain occurs. The notion of *template* denotes a molecular “mold” that dictates the structure of another molecule; most commonly, one strand of DNA that directs synthesis of a complementary DNA strand during DNA replication of an RNA during transcription.

**Fig. 7** Supercoiling process in a double-stranded closed-circular DNA, here pictured in two spatial shapes: the one (left) twisted, the other (right) untwisted, with the respective linking number



in resolving (untangling) finished chromosomes after DNA replication (Summers 1990). Both replicated circular and linear DNA are separated by type II topoisomerases. Type IV topoisomerases comprise two subunits, the ParC and ParE proteins, which are necessary for proper chromosome partition in bacteria. In the early 1990s, it was discovered that these subunits together constituted a type II topoisomerase. The catalytic properties of topoisomerase IV can be distinguished from those of DNA gyrase,<sup>8</sup> which belong to type IIA topoisomerase, in two important ways. First, although topoisomerase IV can remove positive and negative superhelical twists from DNA, it cannot actively underwind the double helix. Second, the ability of topoisomerases IV to resolve DNA knots and tangles is dramatically better than that of DNA gyrase. Because of these differences, the physiological roles of topoisomerase IV are distinct from those of DNA gyrase. The primary cellular functions of topoisomerase IV are to unlink daughter chromosomes following DNA replication and to resolve DNA knots that are formed during recombination. Recently, it was found that topoisomerase IV removes positive supercoils from DNA more efficiently

<sup>8</sup> DNA gyrase is an essential bacterial enzyme that catalyzes the ATP-dependent negative supercoiling of double-stranded closed-circular DNA. Discovered in 1976, gyrase belongs to a class of enzymes known as topoisomerase of type IIA that are involved in the control of topological transitions of DNA. In contrast to other types II topoisomerases, DNA gyrase is the only enzyme that is capable of actively underwinding (i.e., negatively supercoiling) the double helix. It accomplishes underwinding by wrapping DNA around itself in a right-handed fashion (creating thus a positive supercoil) and carrying out its strand passage reaction in a unidirectional manner (thus converting a positive to a negative supercoil). The ability of gyrase to wrap DNA during its strand passage reaction allows it to remove positive supercoils that accumulate in front of replication forks and transcription complexes even faster than it can introduce negative supercoils into relaxed DNA. In other words, the negative supercoiling activity of DNA gyrase far exceeds the ability of the enzyme to remove knots and tangles from the genetic materials. Therefore, the major physiological roles of DNA gyrase stem directly from its ability to underwind (opening) the double helix. Therefore, gyrase maintains negative supercoiling of the genome, facilitating the initiation of transcription and replication. It also relaxes positive supercoils in front of elongating polymerases.

than it removes negative supercoils. This has led to speculation that the enzyme also may act ahead of DNA tracking systems to alleviate overwinding of the double helix. (For a more detailed and comprehensive discussion of the topological functions of topoisomerases, we refer to Bates and Maxwell 2005; Boi 2011a, b; Sutormin et al. 2021).

As it has been underlined, “There is now strong evidence that the class of enzymes known as DNA topoisomerases, which catalyze the breakage and rejoining of DNA strands by two successive transesterifications reactions, are nature’s tools for solving the topological problems of DNA replication (...). Because the topological problems of DNA are deeply rooted in its structure, they surface in many other processes involving DNA. As a consequence, the DNA topoisomerases are involved in nearly all biological transactions of DNA. Recent studies in both prokaryotes and eukaryotes have shown, for example, that these enzymes are involved in the relaxation of negatively and positively supercoiled domains that are generated in a DNA template during transcription. Additional examples are the involvement of eukaryotic DNA topoisomerase II in chromosomal condensation and decondensation, and the involvement of prokaryotic topoisomerases in the regulation of the supercoiled state of intracellular DNA.” (J.C Wang, P.R. Caron, R.A. Kim, 1990, 403).

$$\begin{aligned} \text{Lk}(C_1, C_2) &= 1 & \text{Lk}(C_1, C_2) &= -1 \\ \text{Wr}(B) &= 0, \text{Tw}(B) &= 1 & \text{Wr}(B) = -1, \text{Tw}(B) = 0. \end{aligned}$$

## Topological compaction and DNA supercoiling

One of the most striking phenomena that reveals the profound interdependence between topological problems and biological processes is that of the compaction of the chromatin within the cell nucleus. Its explanation is very challenging both for mathematics and biology. Here, we are faced with a genuine problem of differential topology. *What kind of deformations does the double-stranded linear*

*DNA molecule undergo in order that it condenses into an extremely compact form, corresponding to the metaphase of the chromosome?*

One important aspect concern *supercoiling*, which plays an essential role in biological processes, especially in the unwinding of DNA (important for transcription), in DNA replication (with a reduction of the linking number of DNA molecule), and in condensing DNA and promoting the disentanglement of topological domains.

We have three interrelated mathematical and biological (theoretical and experimental) facts which we would like to stress. (1) DNA condensation is a driving force for double helix unlinking and chromosome portioning, by folding, in topological domains. (2) Condensation is achieved by supercoiling, which is a topological state of macromolecules enhanced by three kinds of deformations (embeddings), twisting, writhing, and knotting. We can define the twist of a ribbon abstractly as the integral of the incremental twist of the ribbon about the axis; so, it simply measures how much the ribbon twists about the axis from the frame of reference of the axis (it need not to be an integer). The writhe measures how much the axis of the ribbon is contorted in space. (3) Supercoiling results from topological strain and the contortion of DNA by proteins.

Here, it can be useful to introduce a related notion (Zee-man 1965; Elhamdadi et al. 2020). A *framed knot*<sup>9</sup> is the extension of a *tame knot* to an embedding of the solid torus  $D^2 \times S^1$  in  $S^3$ . The *framing* of the knot is the linking number of the torus with the knot, i.e., the number of times that the knot intersects the torus.<sup>10</sup> A framed knot can be seen as the embedded ribbon and the framing is the (signed) number of twists. This definition generalizes to an analogous one for

<sup>9</sup> The general definition is as follows. A *framed knot*  $(K, V)$  in  $S^3$  is a knot  $K$  equipped with a continuous non-vanishing vector field  $V$  normal to the knot, called a *framing*. Similarly, a framed link in  $S^3$  is a link  $L$  where each component is equipped with a framing. A framed knot can be visualized as a tangled ribbon that has had its two ends glued after an even number of half-twists, so as to yield an orientable surface. Note that this means we exclude the cases in which the ribbon is glued together after an odd number of half-twists, i.e., a Möbius band. More precisely, the ribbon forms an embedded annulus, one of whose boundary components are identified with the specified knot  $K$ . For a given knot  $K$ , two framings on  $K$  are considered to be equivalent if one can be transformed into the other by a smooth deformation. This is indeed an equivalence relation on the set of framings, and as such, the term “framing” will be used to refer to either an equivalence class or a representative vector field.

<sup>10</sup> We can also give the following definition. Given a knot  $K$  in the 3-sphere  $S^3$ , consider a singular disk  $D^2$  bounded by  $K$  and the intersections of  $K$  with the interior of the disk. The absolute number of intersections defines the *framing* function of the knot. One can show that the framing function is symmetric except at a finite number of points. The symmetric axis is a new knot invariant, called the *natural framing of the knot*. More formally: Let  $K: S^1 \rightarrow S^3$  be an unoriented knot. Let  $D$  be the 2-disk. We define a compressing disk of  $K$  to be the map  $f: D \rightarrow S^3$ , such that  $f|_{\partial D} = K$  and such that  $f|_{\text{int}(D)}$  is transverse

framed links. Framed links are said to be equivalent if their extensions to solid tori are ambient isotopic. Framed link diagrams are link diagrams with each component marked, to indicate framing, by an integer representing a slope with respect to the meridian and preferred longitude. Given a knot, one can define infinitely many framings on it. Suppose that we are given a knot with a fixed framing. One may obtain a new framing from the existing one by cutting a ribbon and twisting it an integer multiple of  $2\pi$  around the knot and glue back again where the cut was made. In this way, one obtains a new framing from an old one, up to the equivalence relation for framed knots, leaving the knot fixed. The framing in this sense is associated with the number of twists the vector field performs around the knot. Knowing how many times the vector field is twisted around the knot allows one to determine the vector field up to diffeomorphism, and the equivalence class of the framing is determined by this integer called the *framing integer*. If we apply the *Kirby calculus*, in which the desired equivalence class of knot diagrams is not a knot but a *framed link*, one must replace the type I move with a “modified type I” move composed of two types I moves of opposite sense. The new type I’ move affects neither the framing of the link nor the writhe of the overall knot diagram. Kirby calculus is a method for modifying framed links in the 3-sphere using a finite set of moves, the Kirby moves. Using four-dimensional Cerf theory, Kirby proved that if  $M$  and  $N$  are 3-manifolds, resulting from Dehn surgery on framed links  $L$  and  $L'$ , respectively, then they are homeomorphic if and only if  $L$  and  $L'$  are related by a sequence of Kirby moves (Kirby 1978). According to the Likorish–Wallace theorem, any closed orientable 3-manifold is obtained by such a surgery on some link in the 3-sphere (Culler et al. 1987).

Supercoiling is a key vector of biological functionality. It is one of the three fundamental aspects of DNA compaction; the other two are conformational flexibility and intrinsic DNA curvature. For example, the problem of DNA compaction in *E. coli* can be putted in the following words (Lal et al. 2016): the DNA must be compacted more than a thousand-fold in the cell, yet it still needs to be available to be transcribed. (Recall that the length of a typical bacterial operon—usually about three genes—is about as long as the entire bacterial cell, if it is stretched out in its B-DNA double-helical conformation!.) In order for compaction to be achieved, some kind of anisotropic flexibility or ‘bendability’ of DNA, which is very much sequence-specific, and is different from the structural ‘rigidity’ of DNA, is required.

Footnote 10 (continued)

to  $K$ . Then,  $f|_{\text{int}(D)}$  has only finitely many intersections with the knot. We call the intersections points the *holes* of the compressing disk, and denote their number by  $n(f)$ . So,  $n(f) = |\{f^{-1}(K) \cap \text{int}(D)\}|$ . The *knottedness* or *linking coefficient*  $Lk(K) = \min\{n(f) \mid f(D) \text{ a compressing disk}\}$  is a basic invariant of the knot  $K$ .

Whereas the *persistence length* of DNA is relatively non-specific, and just has to do with its overall ‘rigidity’ (on average, DNA has a persistence length of about 44 nm, which is quite a bit longer than proteins—one way to think about this is that proteins tend to fold up into little spheres, or ‘blobs’, and DNA is a bit more rigid), *anisotropic flexibility* is a measure of a particular sequence to be deformed by a protein (or some other external forces). Some sequences are both isotropically flexible and ‘bendable’—for example, the TATA motifs (see Venkata and Bansal 2017).<sup>11</sup> Perhaps, one of the best examples of this is the binding site for the Integration Host Factor (IHF): there are certain base pairs that are highly distorted upon binding of this protein. It is quite impressive that this protein induces a bend of 180 degrees into a DNA helix. In other words, the curvature, say  $K$ , at each sequence of the two strands of DNA helix must be very sharp in order that the DNA double helix may assume its extremely compact form. Therefore, the relationship between (geometric) curvature and conformational (or topological) flexibility appear to be crucial in the understanding of the biological activity of cells (see Boi 2007a, b, c).

The DNA molecule is condensed by the action of proteins histones. Indeed, when one considers that the DNA must be compacted more than a thousand-fold in the cell, it is probably not surprising that almost any protein that binds to DNA will bend it. Moreover, since the total curvature  $K$  of an entire DNA double-helix segment depends on the torsional stress which applies to DNA strands,<sup>12</sup> and, accordingly, these strands form a twisted curve, i.e., a curve of double curvature in the three-dimensional space of the cell nucleus, DNA double-helix must coil many times in a very ordered way to form chromatin structure; otherwise, if the chromosome of a human cell were in the form of a random

coil, they would not fit within the nucleus. The DNA double-helix coils first by overwinding or underwinding of the duplex. The supercoiled form of a circular DNA molecule is much more compact than the other possible conformations, i.e., nicked and linear.<sup>13</sup> In its supercoiled form, DNA molecule minimizes to the highest the space volume it occupies in the nucleus. Supercoils condense DNA and promote the disentanglement of topological domains.

Let us remark that there is some significant analogy between the shape and structure of DNA double-helix and the form of a special class of surfaces, namely *minimal surfaces*.<sup>14</sup> It is an object that change as we change the *moduli* (a family of parameters) along the curve, and this may trigger some variation of the morphology of the molecule, which is an important promoter of the functionality of all complex living organisms. Recall that minimal surfaces come in one-parameter families (so-called *associate families*), all of whose members are isometric, though usually not congruent. Using the associated family parameter as a morphing parameter provides a particularly beautiful moving picture or simulation, one that in principle may serve as a simplified model of the real moves of the double helix in the nucleus of the cell. The *helicoid* and the *catenoid* belong to an associate family, and differential geometry books often show several frames of a morph between them (Fig. 8).

In this context, concerning the shape of the DNA molecule and its variations, another very promising line of inquiry at the intersection of mathematics and biology, would be the moduli spaces of higher genus Riemann surfaces. Let us give only few hints and review briefly some recent results. As it is well known to mathematicians, classification problems in algebraic geometry and other parts of geometry often include two steps. The first step is to find as to many discrete invariants as possible (for example, if we want to classify compact Riemann surfaces, then the

<sup>11</sup> In eukaryotes, genes can be broadly classified as TATA-containing and TATA-less based on the presence or absence of a TATA box in their promoter sequences. They have been studied in depth in yeast, and it is reported that TATA-containing genes are expressed at extremely high or low levels, are stress-induced, and are under evolutionary selective pressure, when compared to TATA-less genes. The two classes of genes also vary in their usage of transcription factors (SAGA vs. TFIID) in yeast. Furthermore, in yeast, TATA-containing genes prefer sub-telomeric location in the genome and have more duplicates. The structural features of TATA-containing TATA-less promoters are distinctly different in lower eukaryotes. The TATA-containing core promoters are less stable, more flexible, and more curved compared to TATA-less promoters in *S. cerevisiae*, *C. elegans*, and *D. melanogaster*. In mouse and human, stability and curvature are distinguishing features of TATA-containing and TATA-less promoters.

<sup>12</sup> Chromosomal and plasmid DNA molecules in bacterial cells are maintained under torsional tension and are therefore supercoiled. With the exception of extreme thermophiles, supercoiling has a negative sign, which means that the torsional tension diminishes the DNA helicity and facilitates strand separation.

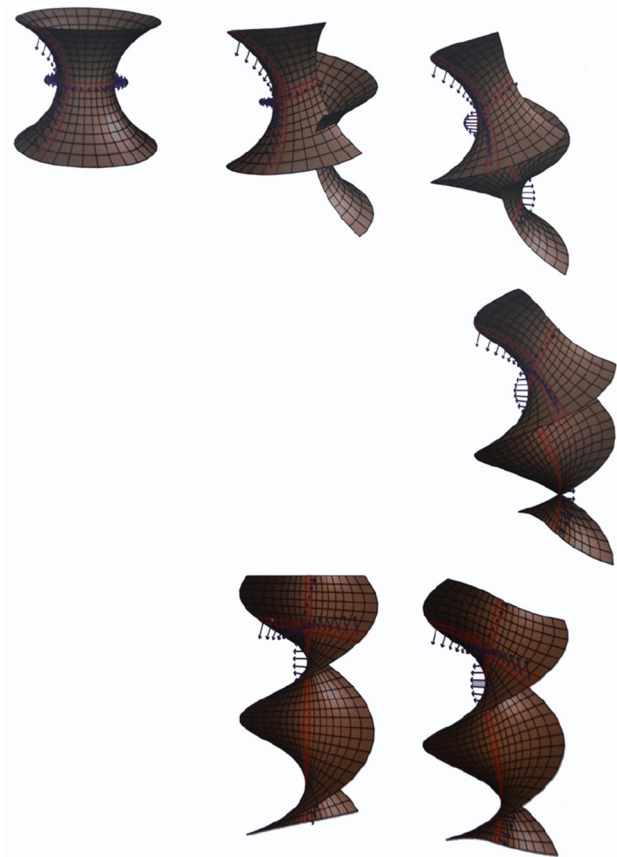
<sup>13</sup> Linear DNA generally migrates between the nicked circle and the supercoiled forms. However, it may also migrate the same distance as nicked circle—it migrates as predicted by the length of the DNA.

<sup>14</sup> Historically, the theory of minimal surfaces was born with the optimization problem formulated by Lagrange: «Given a closed curve in tridimensional space, we have to find that surface which minimizes the area, among all those that have as boundary such a curve». In the 1850's Plateau was the first to understand that each closed curve may be the boundary of a minimal surface. The conjecture, known as the *Plateau's problem*, attracted many mathematicians, and the complete solution is due to Jesse Douglas in 1931 (Douglas 1931). The catenoid is a rotational surface bounded by two circles placed in two parallel planes. It was the first minimal surface known, which was discovered by Euler in 1744; the helicoid was discovered by Lagrange in 1766. The minimal surface has equal surface tension in all their points, which means geometrically that the average curvature  $H$  is  $=0$ . Hence, a minimal surface has, in every point, average curvature  $H=0$ . Such a minimal surface needs not be minimizing for the area.

principal discrete invariant is the genus). The second step is to fix values of the discrete invariants and to try to construct a moduli space; that is, an algebraic variety (or other appropriate space in other parts of geometry) whose points correspond to the equivalence classes of the objects to be classified in some natural way. In a general significance, a moduli space is *the variety of possibilities that a space has to be deformed*; in other terms, *all the shapes that this space may take, up to equivalence*. The mathematical translation of this statement requires a deep analysis and precise definitions of some fundamental concepts such as those of continuous and discrete, local and global, genericity, and singular. Of course, we will not address this study here.

Roughly, a moduli spaces problem consists of three ingredients. Objects: which geometric objects would we like to describe, or parametrize? Equivalences: when we identify two of our objects as being isomorphic, or “the same”? Families: how do we allow our objects to vary, or modulate? The questions that arise naturally are: What these ingredients signify? And what it means to solve a moduli problem? First, let us recall that moduli spaces arise throughout algebraic geometry, differential geometry, and algebraic topology. The basic idea is to give a geometric structure to the totality of objects we are trying to classify. If we can understand this geometric structure, then we obtain powerful insights into the geometry of the objects themselves. Furthermore, moduli spaces are rich geometric objects in their own right. They are meaningful spaces, in that any statement about their geometry has a “modular” interpretation in terms of the original problem. As a result, when one investigates them, one can often reach much further than one can with other spaces.

Let us remark that examples of moduli spaces include two families of key importance, namely, the Riemann moduli space of an orientable topological surface  $S$ , and the moduli space of flat  $G$ -connections on such a surface  $S$ , where  $G$  is some fixed Lie group. The important point here is that the former admits an elementary combinatorial description in terms of “fatgraphs” (discrete topological objects) and was applied to study the topology of RNA (see Penner and Waterman 1993; Bon et al. 2008), and the latter in terms of “ $G$  graph connections” and was used to analyze the geometry of proteins for  $G=SO(3)$ , the group of rigid rotations of 3-space  $\mathbb{R}^3$ . Specifically,  $G$  graph connections allow to probe the geometry of proteins, namely the geometry of hydrogen bonds among peptide units in a protein. The result found by Penner and coworkers is that the rotations cluster into only about 30% of the volume of  $SO(3)$ , and moreover, within this region, there is a further aggregation into 30 sub-regions of clusters (Penner and Waterman). This gives a new classification for the geometry of hydrogen bonding



**Fig. 8** The picture shows how one passes by a series of transformations from a catenoid to a helicoid. The *Bonnet transformation* allows to transform a catenoid in a helicoid. Mathematically, the Bonnet transformation can be described as the weighted sum of two minimal surfaces:  $S = \cos\theta S'' + \sin\theta S'$ , where  $\theta$  is the Bonnet angle (parameter). To biologists, the Bonnet transformation is an attractive way to describe complex reorganizations, which influence the physiological functions of macromolecules like DNA and proteins. Many macromolecules, indeed, attain minimal surface shape, and for the kind of them that undergo dynamic structural changes, the Bonnet transformation result to be very (physiologically and evolutionary) advantageous. This is notably because: there will be a well-defined, low energy path from one state to another; the transformation is isometric, and therefore, no bonds are stretched; it will proceed along a well-defined path, since the isometry is unique; and, perhaps most important in the case of bio-molecules, the parallel surface undergoes a Bonnet transformation as well, leaving any hydration shell virtually unperturbed. Bonnet transformation is a very interesting example of a geometric structure which optimizes the biological processing and work of living matter

that unifies and extends those already known. For RNA, it is not the geometry but rather the topology which is useful for describing its structure. Penner showed that, in fact, there is a natural decomposition of the Riemann moduli space for a surface  $S$  whose cells are in one-to-one correspondence with homotopy classes of suitable graphs embedded in  $S$ .

There is, moreover, a natural combinatorial model based on chord diagrams<sup>15</sup> for the moduli space of  $r$  interacting molecules, which have a genus  $g$  in a suitable sense. The striking theorem established by Penner and coworkers is that the Riemann moduli space of a surface  $S$  of genus  $g$  with  $r$  boundary components is combinatorially isomorphic with this RNA moduli space up to homotopy (see Andersen et al. 2013; Penner 2016).

## A topological approach to the study of biological processes

The study of some processes of biological systems can be addressed through differential geometry and topological knot theory (Boi 2005), which allows for modeling the three-dimensional structures of DNA and protein–DNA complexes (Boi 2021). The difficult task is first to show that certain topological deformations associated to the supra-molecular structures during the cell cycle take part in the dynamics of chromatin, the organization of chromosome, and also in the cell’s activities. And then to elucidate the way in which these deformations might modulate the action of different regulatory systems, ensuring in particular the transition of this action from a local-target mechanism to global functional processes.

We will focus on three work hypotheses. First, we argue that the interaction between topological changes and dynamical processes constitute a deep and largely unexplored meeting point for mathematics and biology. Then, we assume that certain geometric properties and topological patterns work like dynamical principles, meaning that they are intrinsically involved in the organization and growth of living systems. Finally, we claim that these properties and patterns display intricate biological plasticity and complexity on every scale, from the very large (organism) to the very small (molecule).

Let us start with, say, the “basic” level of DNA structure and chromatin dynamics. We will describe some aspects of the way in which (1) the two strands of DNA must be continuously unlinked during replication, and (2) the chromatin is topologically condensed within the cells of organisms with nuclei. There are three families of huge ATP-powered enzymes, helicases, type II topoisomerases, and condensins, which contribute to the orderly unlinking of DNA and to the chromosome segregation *in vivo*. This twofold process seem

to be really fundamental for the fate of our organism. For replication to occur, the DNA must initially be decondensed. Helicases unwind DNA creating (+) supercoiling and precatenanes, which are rapidly removed by topoisomerases.<sup>16</sup> Type-2 topoisomerases actively remove all DNA entanglements (Cozzarelli et al. 1985). Then, the organized recombination by condensins and supercoiling are essential for chromosome partitioning. The chromosome must, indeed, be folded into topological domains. Besides, chromosome needs to be topologically remodeled in order that the genetic events and cellular processes may be performed.

Let us briefly explain what is meant by *supercoiling* (see below for a thorough description of this concept). The supercoiling of a closed-circular molecule into an interwound superhelix can be understood in terms of the relation between three mathematical quantities: *linking*, *writhe*, and *twist*. From the very onset, it is important to stress the fact that the topological state of DNA and the level of its supercoiling can be explained using the linking number concept (Lk). In the case of a covalently closed-circular double-stranded DNA molecule, its linking number is the number of intersections of one strand with the second strand, with allowance for the sign of this intersection. The linking number Lk does not depend on the molecule deformations and can only be altered through cleavage, passage, and relegation of DNA strands. It is hence a topological invariant obtained by adding the two geometrical parameters, namely the twist and the writhe. The *twist* is defined as the number of time DNA chains turns around each other around the double-helix axis, while the *writhe* is a measure of the supercoiling of the DNA axis. In nature, supercoiled DNA in the form of writhe stably exists in two forms: plectoneme (a higher order double helix) and a solenoid (a higher order single helix) which is typical of DNA wrapped around a protein. An interesting solenoid model of the chromosomes, that is of the wrapping of the DNA–histone proteins into the chromatin, was proposed by the French biologist Képès and Vaillant (2003). The main idea was that an ordered solenoidal supercoiled organization would facilitate the co-expression of groups of genes. (For a more detailed and comprehensive discussion of DNA topology, see Bates and Maxwell 2005; Boi 2011a, b; Jost et al. 2014).

Because molecule is underwound (DNA string with fewer than the natural number of turns is known as *underwound*, while with more than the natural number of turns is known as *overwound*; DNA molecule supercoils in Nature are usually underwound rather than overwound), it has a deficit in

<sup>15</sup> A chord diagram is a finite trivalent undirected graph with an embedded oriented circle and all vertices on that circle, regarded modulo cyclic identification, if any. Equivalently, this is a pairing (by chords) of all elements in a cyclic order (the boundary vertices). Topologically, a chord diagram is an even number of distinct points on the circle, grouped in pairs, up to an orientation preserving homeomorphism of the circle. Such a diagram is pictured by a certain number of chords with distinct endpoints in a circle.

<sup>16</sup> Stated differently, unwinding of the helix during DNA replication (by the action of helicase) results in supercoiling of the DNA ahead of the replication fork. This supercoiling increases with the progression of the replication fork. If the replication supercoiling is not relieved, it will physically prevent the movement of helicase.



linking number compared with a relaxed molecule of the same size. It compensates by writhing and by twisting and bending, satisfying the equation  $Wr = Lk - Tw$ . Furthermore, the linking number of course is given by the total amount of writhes plus the total amount of twists:  $Lk = Wr + Tw$ . The remarkable fact about this result is that two geometric quantities (writhing and twisting) may change under deformations of the curve sum to a topological quantity (linking number), which is invariant under such deformations. The linking number of DNA double helix in all organisms is less than the energetically most stable value in unconstrained (relaxed) DNA. This puts DNA under (physical) stress which causes it to buckle and coil in a regular way called (–) supercoiling. The (–) sign indicates that the linking number is less than in the relaxed state. The name supercoiling arises, because it is the coiling of a molecule which is itself formed by the many-times coiling of two strands about each other. Although supercoiling is, strictly speaking, a geometric property, it is a consequence of a topological one, the linking number difference between supercoiled and relaxed DNA.

It clearly follows from the previous facts that DNA in living systems is topologically constrained; precisely, its geometric (local) structure depends on how it is topologically (globally) constrained. Organisms are faced with two main mathematical problems: (1) the unlinking of DNA during replication, and (2) the partitioning of the chromosome. Topoisomerases actively remove all DNA entanglements. More precisely, the topological constraints on DNA generally involve the regulation of its *linking number* by the transient cutting by enzymes. In other words, topoisomerases are enzymes that participate in the overwinding or underwinding of DNA. Underwinding DNA facilitates a number of structural changes in the molecule. Strand separation occurs more readily in underwound DNA. This is critical to the processes of replication and transcription, and represents a major reason why DNA is maintained in an underwound state.

To prevent and correct these types of topological problems caused by the double helix, topoisomerases bind to DNA and cut the phosphate backbone of either one or both the DNA strands. The activity of DNA, including gene expression and replication, depends sensitively on the linking number imposed, which is a topological invariant. This topological invariant can be decomposed into the sum of two geometric invariants, the *twisting* and the *writhing*, whose analysis involves integral geometry (see below for more details).

The double-helix DNA is a multifaceted spatial structure. It is both a geometrical entity and a topological form. This topological form is itself a manifestation of linking and knotting. Within the cell, the DNA is a very long molecule with a remarkably complex topology. Topological properties of DNA are defined as those that can be changed only by breakage and reunion of the backbone, that is by surgery (cut and

gluing). As we will see, this complex topology of DNA is essential for the life of organisms.

The topology of DNA in vivo is set by a remarkable group of enzymes called *topoisomerases*. As we already mentioned, these enzymes essentially promote the passage of DNA segments through each other until a stable state is achieved. This functional stability is thus made possible thanks to a conformational/topological flexibility of the double-helix, and the continuous remodeling of nuclear structures is as well required for cell activity to be performed. There are three important topological properties of DNA:

1. The linking number between two strands of the double helix.
2. The interlocking of separate DNA rings into what are called *catenanes*.
3. Knotting. We will return shortly on these properties with more details.

Likewise, we observe three physical and phenomenological properties of the molecule, which can be briefly described as follows:

- (a) As the number of crossing in a knot or catenane increases, the number of possible isomers grows exponentially;
- (b) The linking number of DNA in all organisms is less than the energetically most stable value in unconstrained (relaxed) DNA; this puts the DNA under stress, which causes it to buckle and coil in a regular way called negative (–) supercoiling.
- (c) The name supercoiling arises, because it is the coiling of a molecule, which is itself formed by the coiling of two strands about each other. Although supercoiling is, strictly speaking, a geometric property, it is a consequence of a topological one, the linking number difference between supercoiled and relaxed DNA.

The stable structures of the DNA molecule are those that minimize a conformational energy subject to constancy of the topological conditions. This fact gives rise to a range of variational problems. Experiments show that the stable structures of proteins minimize energy. Let us stress that the native structure of a protein is the thermodynamically stable one, as showed by Anfinsen's experiments<sup>17</sup> (see

<sup>17</sup> Anfinsen's experiments concern protein folding. In the 1950s, Christian Anfinsen conducted a series of experiments in which he determined that all the information needed to form the three-dimensional structure of the protein (polypeptide chain) is stored in the specific sequence of amino acids in that polypeptide. Later experiments confirmed this fact, i.e., that primary structure determines the final confirmation of the protein. In his first experiment, Anfinsen used some appropriate denaturing agents to break down the

Anfinsen 1973); and also, that although a protein's folding pathway(s) can depend sensitively on sequence, there are proteins, described quite accurately by energetically non-frustrated models, where the topography of the free energy is determined just by native topology. Thus, to predict protein structures from sequences, one must solve an optimization problem. Disagreeing with one of the central tenets of molecular biology, which states that globular proteins have a unique 3-D structure or fold that fosters its function (Anfinsen's postulate), recent work has identified several fold-switching proteins whose secondary structures can be remodeled in response to a few mutations (evolved fold switchers) or cellular stimuli (extant fold switchers) (Porter and Looger 2018). Another aspect of protein folding is essentially topological. In the last 2 decades, several studies of solved protein structures have demonstrated the existence of many deeply knotted proteins. Conservation of knotting across some protein families strongly suggests that knotting can be important for protein structure and function; it hence appears significant to understand how protein knots forms and in which specific physiological contexts they form. More recent work investigating the folding and unfolding of the slip-knotted archaeal virus protein AFV3-109 revealed that the unfolding of this protein proceeds through a folding intermediate that has the topology of a trefoil knot. Furthermore, the rate of slip-knot formation rapidly increases either when one increases the relative stiffness of bending, or when one decreases the speed of ambient coiling. (see Begun et al. 2021).

Footnote 17 (continued)

secondary and tertiary structure of ribonuclease. Precisely, he used urea agent to break down non-covalent bonds (also called *disulfide bounds*) such as hydrogen bonds holding the secondary structure, and then, he used the beta-mercaptoethanol to reduce and break down the disulfide bonds holding the tertiary structure together. The effect of the exposition of the native enzyme to these two agents was the complete denaturation of the protein. And when he removed the two agents simultaneously via dialysis, he found that the protein refolded back into its original biological active form. Then, in a second experiment, instead of removing the two agents at the same time, he first removed the beta-mercaptoethanol, and afterward, he removed the urea. What Anfinsen discovered was that the final protein refolded but became scrambled and was no longer biologically active. The hypothesis putted forward by Anfinsen was that this happened, because the non-covalent bonds could not form in the presence of urea, and so, disulfide bonds formed incorrectly. In a third experiment, he found that if he exposed the scrambled, inactive protein to trace amounts of beta-mercaptoethanol in the absence of urea, the biologically active native structure eventually reformed. This happens, because the tiny amount of beta-mercaptoethanol was enough to catalyze the breaking of the incorrect disulfide bonds. Finally, the protein formed the correct disulfide bridges and returned to its native form, because this was thermodynamically most stable and lowest in energy form.

## Some topological concepts

Recall briefly that, mathematically, a *knot*  $K$  is an embedding of a one-dimensional closed curve into  $S^3$  or  $\mathbb{R}^3$ . A *link*  $L$  of  $m$  components is a subset of  $S^3$ , or of  $\mathbb{R}^3$ , that consists of  $m$  disjoint, simple closed curves. A link of a component is a knot. To establish the *equivalence* between links, we need the topological notion of *homeomorphism*. Then, we can state that links  $L_1$  and  $L_2$  in  $S^3$  are equivalent if there is an orientation-preserving homeomorphism  $h: S^3 \rightarrow S^3$ , such that  $h(L_1) = L_2$ .

The analytical formula for the *linking number* of a pair of entangled curves is

$$Lk(\gamma_1, \gamma_2) = 1/4\pi \int_{\gamma_1 \times \gamma_2} dx/ds \times dy/dt \cdot (x-y)/|x-y|^3 ds dt. \quad (1)$$

The linking number of a pair of knotted curves is a numerical invariant (an integer number). It is an invariant under *Reidemeister moves*, which means that when we move slightly and smoothly any part of the diagram, the linking number does not change. Any two diagrams of equivalent links  $L_1$  and  $L_2$  are related by a sequence of *Reidemeister moves* and an orientation-preserving homeomorphism of the plane. A *link diagram* of  $L$  is the image of  $L$  in  $\mathbb{R}^2$  together with 'over and under' information at the crossing. A crossing is a point of intersection of the projections of two-line segments of  $L$ . The Reidemeister moves are of three types and each replaces a simple configuration of arcs and crossings in a disc by another configuration:

1. Twist and untwist in either direction (a rigorous definition of *twist* and *writhe* was given in section "[Geometry of the DNA: the linking number and its connection with genomic processes](#)" and).
2. Move one loop completely over another.
3. Move a string completely over or under a crossing.

The type I move is the only move that affects the writhe of the diagram. The type III is the only one that preserves the number of crossings of the diagram. Any homeomorphism of the plane must preserve all crossing information. In other words, and following a theorem by Reidemeister (1927), all changes of knot or link diagrams can be obtained by performing three basic motions applied just to small portions of the diagrams near the crossings, along with simple deformations in the plane, called *plane isotopies*, which do not change any of the crossings of diagrams.

To specify the notion of *isotopy*, let us give the following definition (Kauffman 1990): there exist  $h_t: S^3 \rightarrow S^3$  for  $t \in [0, 1]$ , so that  $h_0 = 1$  and  $h_1 = h$  and  $(x, t) \mapsto (h_t(x), t)$  is a piecewise linear homeomorphism of  $S^3 \times [0, 1]$  to itself. In

this way, the whole of  $S^3$  can be continuously deformed, using the homeomorphism  $h_t$ , at time  $t$ , to move  $L_1$  to  $L_2$ . A link or a knot invariant may be thought of as a quantity that remains unchanged when we apply any one of the previous Reidemeister moves to a regular diagram. Moreover, if one link diagram of an oriented link is changed into another diagram for an oriented link by any Reidemeister move, the linking number does not change. This is true in the special cases of moves type I and type II. Thus, we have the important results that *the linking number is an invariant of unoriented two-component links*. Precisely, there is a theorem which states that *if two equivalent (unoriented) links of two components are each oriented in any way, then the absolute value of their linking numbers will be equal*.

Let us now return to the geometric structure of DNA. An important general point that needs to be stressed is that the topological deformability of the DNA molecule, the structural modifications of the chromatin and the spatial architecture of chromosome exert an important influence on the way in which DNA acts within cell. The remodelers (i.e., families of proteins' regulatory complexes) of chromatin structure play a fundamental role in replication and repair of DNA sequences and in the transcriptional activities of the entire genome.

We must consider the basic level of the DNA structure which is its coiling, and then try to understand the mechanisms responsible for the knotting and unknotting of the double-helix. Large amounts of DNA molecule are wound up and packed into the average cell. DNA molecule is an incredibly long polymer, whereas the cell's nucleus has a very thin spatial volume. This obviously means that the embedding of DNA into chromatin within the cell core is exceedingly complicated. Therefore, many complex structural modifications, topological deformations, and regulatory network interactions must work together to perform the proper packing of DNA into several folding levels of chromatin.

We suggest that there must be a deep connection between topological knot theory and molecular biology, and that knotting and unknotting are 'universal' scale-invariant operations acting on condensed and living matter phenomena, and this should lead us to postulate some significant analogies between the macroscopic, mesoscopic, and microscopic scales and levels of organization of matter. This claim rests principally upon the three following considerations:

1. The spatial conformation of DNA knots is a phenomenon involved in almost all fundamental genetic events.
2. Far from being an accidental fact, we can indeed observe significantly that these molecular knots carry precious information on the emergence of new levels of functionality in living organisms.

3. As a special case of (1), it can be said that some topological contortions of the double-helix molecule, as well as some spatial distortions like bending, twisting, and coiling, carried out by some proteins and enzymes topoisomerases which bind to a large variety of DNA sites, are essential for many biological processes to be performed.

The previous remarks suggest that the geometric transformations and topological deformations associated with many molecular as well as cellular processes during embryogenesis must be considered as an additional layer of biological functionality having real dynamical effects on the global metabolism of living organisms.

Precisely, differential geometry and knot theory can be used to describe three-dimensional structure of DNA and protein–DNA complexes. Biologists devise experiments on circular DNA to elucidate 3-dimensional molecular conformations like helical twist, supercoiling, and the action of various important life-sustaining enzymes such as topoisomerases and recombinases. These experiments are often performed on circular DNA molecules, in which changes in the geometric (curvature, writhing, twisting, and supercoiling) or topological (knotting and linking) state of DNA can be directly observed.

## The White formula and its biological significance

The link between the structure of the DNA double-helix and some differential geometric concepts appears in the *White's formula* relating the linking, twisting, and writhing properties of a space curve. It is useful to start with the "Jordan Curve Theorem" (a mathematical prerequisite of White formula), which states that *A simple, closed, continuous (or smooth, or piecewise linear) curve separates the plane  $\mathbb{R}^2$  into two parts with the property that it is impossible to get from one part to the other by means of a continuous path avoiding the given curve*. The same conclusion holds for any complete curve in  $\mathbb{R}^2$ , i.e., a simple, continuous, unboundedly extended, non-closed curve both of whose ends go off to infinity, without nontrivial limit points in the finite plane.

There is another less obvious generalization of this principle, in three-dimensional space  $\mathbb{R}^3$ . (or in the 3-sphere  $S^3$ ). First consider two continuous (or smooth) simple curves (loops) in  $\mathbb{R}^3$  which do not intersect

$$\begin{aligned} \gamma_1(t) &= (x_{11}(t), x_{21}(t), x_{31}(t)), \gamma_1(t + 2\pi) = \gamma_1(t) \\ \gamma_2(t) &= (x_{12}(t), x_{22}(t), x_{32}(t)), \gamma_2(t + 2\pi) = \gamma_2(t). \end{aligned} \quad (2)$$

Next consider a 'singular disc'  $D_i$  bounded by the curve  $\gamma_i$ , i.e., a continuous map of the unit disc into  $\mathbb{R}^3$ :  $x_i^a(r, a)$ ,

$i = 1, 2, 3$ , where  $0 \leq r \leq 1, 0 \leq a \leq 2\pi$ , sending the boundary of the unity disc onto  $\gamma_i$

$$x_i^a(r, \phi)|_{r=1} = x_i^a(\phi), \quad a = 1, 2, 3, \tag{3}$$

where  $\phi = t$  for  $i = 1$ , and  $\phi = \tau$  for  $i = 2$ . Therefore, we have the following definition: two curves  $\gamma_1$  and  $\gamma_2$  in  $\mathbb{R}^3$  are said to be nontrivially linked if the curve  $\gamma_2$  meets every singular disc  $D_1$  with boundary  $\gamma_1$ , or, equivalently, if the curve  $\gamma_1$  meets every singular disc  $D_2$  with boundary  $\gamma_2$ .

In  $n$ -dimensional space  $\mathbb{R}^n$ , certain pairs of closed surfaces may be linked, namely submanifolds of dimension  $p$  and  $q$  where  $p + q = n - 1$ . In particular, a closed curve in  $\mathbb{R}^2$  may be linked with a pair of points (a ‘zero-dimensional surface’)—this is just the original principle that a simple closed curve separates the plane.

The notion of linking coefficient of two curves was first given by in the 1820s by C. F. Gauss. Specifically, he introduced an invariant of a link consisting of two simple closed curves  $\gamma_1, \gamma_2$  in  $\mathbb{R}^3$ , namely the signed number of turns of one of the curves around the other, the linking coefficient or linking number  $\{\gamma_1, \gamma_2\}$  of the link. His formula for this is

$$N(\text{Lk}) = \{\gamma_1, \gamma_2\} = 1/4\pi \int_{\gamma_1} \int_{\gamma_2} ([d\gamma_1(t), d\gamma_2(t)] \gamma_1 - \gamma_2) / |\gamma_1(t) - \gamma_2(t)|^3, \tag{4}$$

where  $[,]$  denotes the vector (or cross) product of vectors in  $\mathbb{R}^3$  and  $(,)$  the Euclidean scalar product. Thus, this integral always has an integer value  $N$ . If we take one of the curves to be the  $z$ -axis in  $\mathbb{R}^3$  and the other to lie in the  $(x, y)$ -plane, then the previous formula gives the net number of turns of the plane curve around the  $z$ -axis. It is interesting to note that the coefficient  $N$  may be zero, even though the curves are nontrivially linked. Thus, this non-zero value represents only a sufficient condition for nontrivial linkage of the loops.

Now, to explain White’s formula, let  $C$  be a space curve with a unit normal framing  $v, v^\perp$  and unit tangent  $t$  ( $v$  and  $v^\perp$  are perpendicular to each other and to  $t$ , forming a differentiable varying frame,  $\langle v, v^\perp, t$ , at each point of  $C$ .) Let  $C_\epsilon$  be the curve traced out by the tip of  $\epsilon v$  and for  $0 < \epsilon < 1$ . Let  $Lk = Lk(C, C_\epsilon)$  be the linking number of  $C$  with this displacement  $C_\epsilon$ . Define the total twist,  $\text{Tw}$ , of the framed curve  $C$  by the formula

$$\text{Tw} = 1/2\pi \int v^\perp \cdot dv. \tag{5}$$

Given  $(x, y) \in C \times C$ , let  $e(x, y) = (y - x) / |y - x|$  for  $x \neq y$  and note that  $e(x, y) \rightarrow t/|t|$  (for  $t$  the unit tangent vector to  $C$  at  $x$ ) as  $x$  approaches  $y$ . This makes  $e$  well defined on all of  $C \times C$ . Thus, we have  $e: C \times C \rightarrow S^2$ . Let  $d\Sigma$  denote the area element on  $S^2$  and define the (spatial) writhe of the curve by the formula

$$\text{Wr} = 1/4\pi \int_{C \times C} e * d \Sigma = 1/4\pi \int_{z \in S^2} \text{Cr}(z) dz. \tag{6}$$

where  $\text{Cr}(z) = \sum_{p \in e^{-1}(z)} J(p)$  where  $J(p) = \pm 1$  according to the sign of the Jacobian of  $e$ . One can see, from this description, that the writhe coincides with the flat writhe (sum of crossing signs) for a curve that is (like a knot diagram) nearly embedded in a single plane. With these definitions, White’s theorem reads

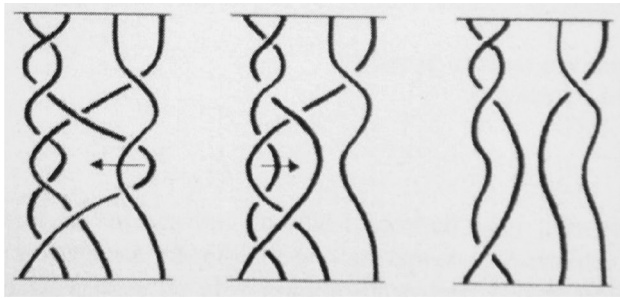
$$\text{Lk} = \text{Tw} + \text{Wr}. \tag{7}$$

This equation is fully valid for differentiable curves in three-dimensional space. Note that the writhe only depends upon the curve itself; it is independent of the framing. By combining two quantities (twist and writhe) that depend upon metric consideration, we obtain the linking number—a topological invariant of the pair  $(C, C_\epsilon)$ . The linking number is a mathematical quantity existing in dimension 3 ( $S^3$  or  $\mathbb{R}^3$ ) for disjoint embedded curves, and in higher dimensions for disjoint embedded closed manifolds (see Kervaire 1965; Rolfsen 1976); a topological invariant by deforma-

tion, which tells us a great deal about the structural properties and qualitative behavior of DNA during the cell cycle. First, it is closely related to the number of time that the two sugar-phosphate chains of DNA wrap around one another. Here, take DNA in its stress-free, relaxed state as the reference point for counting  $\text{Lk}$ , where  $\text{Lk} = 0$ . Now, consider the simple model of a circular DNA with the values:  $\text{Tw} = +3, \text{Wr} = 0, \text{Lk} = +3$ . Thus,  $\text{Lk} = +3$  tells us that DNA has three more double-helical turns than it would have in a relaxed, open-circular form. In general,  $\text{Lk}$  measures the total excess or deficit of double-helix turns in the molecule.

Let  $K$  be a knot, where the word ‘‘knot’’ refers to a representative or to an equivalence class of representatives. (Recall that two knots are equivalent if they are of the same knot type). We will here essentially be concerned with links or knots diagrams of minimal complexity, i.e., ones with the fewest crossings possible. This minimum number of crossings is the crossing number of the link or knot, and a diagram which exhibits the minimum number of crossings is a minimal diagram.

There is an experimental strategy which consists to observe the enzyme-caused changes in the geometry (supercoiling) and the topology (knotting and linking) of the DNA, and to deduce enzyme mechanisms from these changes. This can be called the topological approach to enzymology and is schematically depicted in the following scheme:

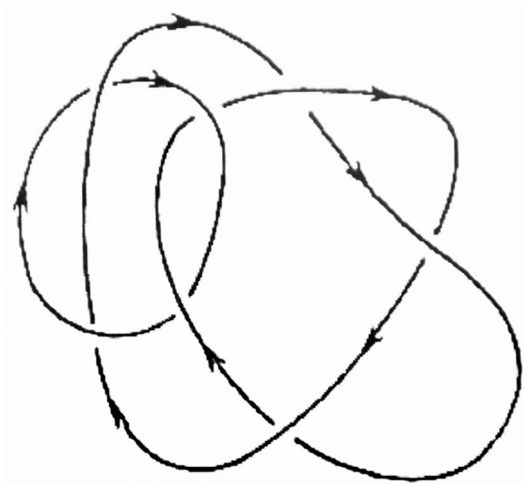


**Fig. 9** Isotopy of a four-string braid (Braid theory is the branch of topology and algebra concerned with braids. A braid of  $n$  strings, denoted  $B_n$ , is an object consisting of two parallel planes  $P_0$  and  $P_1$  in three-dimensional space  $\mathbb{R}^3$ , containing two ordered sets of points  $a_1, \dots, a_n \in P_0$  and  $b_1, \dots, b_n \in P_1$ , and of  $n$  simple non-intersecting arcs  $l_1, \dots, l_n$  intersecting each parallel plane  $P_i$  between  $P_0$  and  $P_1$  and joining the points  $\{a_i\}$  to  $\{b_i\}$ ,  $i=1, \dots, n$ . It is assumed that the  $a_i$ 's lie on a straight line  $L_a$  in  $P_0$  and the  $b_i$ 's on a straight line  $L_b$  in  $P_1$  parallel to  $L_a$ ; moreover,  $b_i$  lies beneath  $a_i$  for each  $i$ . Braids can be represented in the projection on the plane passing through  $L_a$  and  $L_b$ ; this projection can be brought into general position in such a way that there are only finitely many double points, each two of which lie at different levels, and the intersections are transversal.)

*Substrate*  $\rightarrow$  *Reaction*  $\rightarrow$  *Product* (1, supercoiled; 2, knotted; 3, linked) (Figs. 9, 10, 11).

The geometry (supercoiling) and topology (knotting and linking) of circular substrate are experimental control variables. The geometrical and topological properties of the enzyme's reaction products are the observables. In Fig. 12, we start with an unknotted substrate molecule with one negative supercoil. We then show a spectrum of possible products, ranging from an unknotted molecule with 2 negative supercoils (a change in supercoiling) to a trefoil knot (a change in unknotting), to an Hopf link (a change in linking).

A genetic mechanism may engender changes in the genetic code. Site-specific recombination is one of the ways nature geometrically alters the genetic program of an organism, either by moving a block of DNA to another position on the molecule (a move performed by a *transposase*), or by integrating a block of viral DNA into a host genome (a move performed by *integrase*) (Vazques and Summers 2004; Buck and Valencia 2011). An enzyme which mediates site-specific recombination on DNA is called a *recombinase*. A

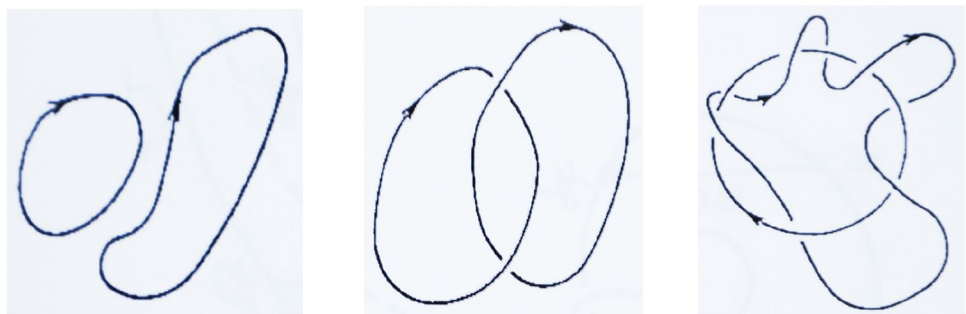


**Fig. 11** The linking coefficient is =0, yet the curves are not trivially linked

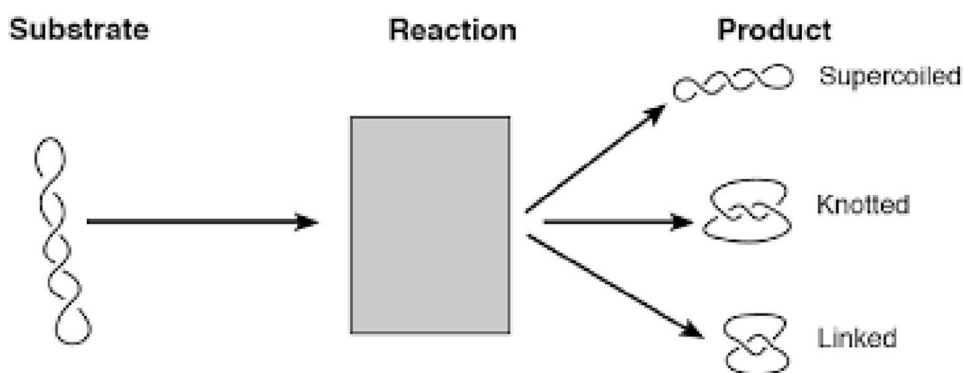
*recombination site* for a given recombinase is a short (10–15 base pairs) linear segment of DNA whose genetic sequence is recognized by the recombinase. Site-specific recombination can occur when a pair of sites (on the same or on different DNA molecules) become juxtaposed in the presence of the recombinase. The pair of recombination sites is aligned (brought close together), probably through enzyme manipulation or random thermal motion (or both), and both sites (and perhaps some contiguous DNA) are then bound by the enzyme (Flapan et al. 2014).

In the recombination event, we have the stage of the reaction which is called *synapsis*, and the term *synaptosome* designates the protein–DNA complex formed by the bound DNA and the enzyme. We will call the entire DNA molecule involved in synapsis (which includes the parts of the DNA molecule not bound to the enzyme) together with the bound enzyme, the *synaptic complex*. After forming the synaptosome, the enzyme then performs two double-stranded breaks at the sites, and recombines the ends by exchanging them in an enzyme-specific manner. The synaptosome then dissociates, and the DNA is released by the enzyme. By analogy with a chemical reaction, we may define a kind of topological reaction and thus call the

**Fig. 10** **a** Unlinked curves. **b** Linking coefficient = 1. **c** Linking coefficient = 4



**Fig. 12** Starting from an unknotted DNA-string substrate, a reaction mediated by an enzyme recombinase may generate a variety of products: supercoiled, knotted, and linked



pre-recombination unbound DNA molecule the *substrate*, and the post-recombination unbound DNA molecule the *product*.

### DNA recombination and the role of mathematical tangles

Let us start this section by giving some basic facts about the biological process of recombination. DNA replication allows for faithfully reproducing the genome from one generation to another. During this process, the correct sequence is maintained by DNA-repair processes throughout the life of a cell and organism. The fundamental process by which the genome can change to generate new combinations of genes is recombination between homologous (or not homologous) DNA sites. Specifically, blocks of genes from homologous chromosomes could be exchanged by the process of crossing-over, or homologous recombination, which takes place during meiosis in sexually reproducing organisms. Recall that each homologous paternal and maternal chromosome contains a different combination of alleles. By generating new chromosomes that contain part of each homologous paternal and maternal chromosome, recombination results in new combinations of alleles on a given chromosome. Thus, recombination provides a mechanism for generating genetic diversity beyond that achieved by the independent segregation of chromosomes.

The events in a reciprocal recombination are equivalent to the breakage of two homologous duplex DNA molecules, an exchange of *both* strands at the break, and a resolution of the two duplexes, so that no tangle remains. The frequency of recombination between two sites is proportional to the distance between the sites. Several types of proteins catalyze steps in recombination.

One of the first models for describing recombination was proposed by Robin Holliday in 1964. After two homologous double-stranded DNA molecules become aligned, a nick is made in one strand of each of the recombining DNAs (step 1). The two nicked strands then invade each other, a process

called *strand exchange*, at the site of the nicks, and the cut 3' ends are joined to the 5' ends of the homologous strand, producing a crossed-strand Holliday structure (step 2). The branch point then migrates, creating a heteroduplex region containing one strand from each parental DNA molecule (step 3).

Rational tangles are not only beautiful mathematical objects but also have many applications in other fields such as biology and DNA synthesis, especially genetic recombination. The theory of tangle was invented in 1986 by J. H. Conway. He introduced the notion of rational tangles, and with each rational tangle, he associated a rational number by the continued fraction method. The associated rational number is based on the pattern of tangle twists. According to Conway's theorem, two rational tangles are equivalent if and only if they represent the same rational number.<sup>18</sup> The classification of rational tangles is crucial for the tangle analysis of site-specific recombination (see Darcy 2014). To each equivalence class of rational tangles corresponds a classifying vector, called the *Conway symbol*. The Conway symbol, an integer entry vector  $(a_1, a_2, \dots, a_m)$ , satisfies the following conditions:  $|a_1| > 1$ ; all entries are non-zero, except possibility for  $a_m$ ; and all entries have the same sign. The classification of rational tangles states that there exists a one-to-one correspondence between equivalence classes of rational tangles and the extended rational numbers  $q/p \in \mathbb{Q} \cup \{\infty\}$  with  $p \in \mathbb{N} \cup \{0\}$ ,  $q \in \mathbb{Z}$  and  $(p, q) = 1$ . Several useful operations can be defined between tangles. Tangle addition shows that: (1) the sum of two rational tangles is not necessarily a rational tangle; it can be a prime tangle. (2) The numerator and denominator operations produce knots and links. (3) The numerator for the sum of two rational tangles is a 4-plat. Every 4-plat can be drawn as a closed braid in four strands, with one untangled strand. (See Vazques and

<sup>18</sup> That is a number that can be expressed as the quotient or fraction  $p/q$  of two integers, a numerator  $p$  and a non-zero denominator  $q$ . Every integer is a rational number, for example,  $5 = 5/1$ .

Summers for a detailed discussion of tangle theory and its relationships biological recombination).

A  $n$ -tangle is a proper embedding of the disjoint union of arcs into a 3-ball; the embedding must send the endpoints of the arcs to  $2n$  marked points on the ball boundary. In mathematical knot theory (Gordon 2006), where a link is defined as a collection of knots which do not intersect, but which may be linked or knotted together (classical examples of links are the Borromean rings, the Hopf link and the torus link), a tangle is an embedding of  $n$  arcs and  $m$  circles into  $\mathbb{R}^2 \times [0, 1]$ ; this definition includes both arcs and circles, and also the possibility of partitioning the boundary of the tangle into two pieces. For example, the  $(-2, 3, 7)$  pretzel knot has two right-handed twists in its first tangle, three left-handed twists in its second, and seven left-handed twists in its third. Analogously to knot theory, we define two  $n$ -tangles as *equivalent* if there is an ambient isotopy<sup>19</sup> (a kind of continuous deformation of the ambient space) of one tangle to the other keeping the boundary of the 3-ball fixed. When we consider a set of marked points on the 3-ball boundary to lie on a great circle, then we may arrange the tangle to be in a general position with respect to the projection onto the flat disc bounded by the great circle. The projection then gives us a tangle diagram with over and under-crossings, as with knot diagrams (see Boi 2021b, c for an in-depth presentation of this subject). From the previous description, we now define a *rational tangle* is a 2-tangle that is isomorphic to the trivial 2-tangle by a map of pairs consisting of the 3-ball and two arcs. We refer, by convention, to the four endpoints of the arcs on the boundary circle of a tangle diagram as being the four directions (or orientations) of the tangle. (we refer to Conway 1970; Ernst and Summers 1990; Kauffman and Lambropoulou 2004, for more details on the topological and algebraic theory of tangles).

It has been stressed that rational tangles and their fractions can be applied to molecular biology (Ernst and Summers 1990; Goldman and Kauffman 1997). “Recombination of DNA is the process of cutting two neighboring strands with an enzyme and then reconnecting them in a different way. The idea of applying tangle theory is to use the addition of tangle to write the equations for possible recombination of DNA molecules. Then one uses topological information (such as the fraction of tangles) to obtain limitations on the possibilities for the products of the recombination. Recombination occurs in successive *rounds* for which the nature of the products can be known through a combination of electrophoresis and electron microscopy. In particular,

electron microscopy provides the biologist with an enhanced image of the DNA molecule from which it is possible to see direct evidence of knotting and supercoiling. In the case of TN3 resolvase, a species of closed-circular DNA is seen to produce very specific knots and links in successive rounds of recombination. By knowing these actual products of the rounds of recombination it is possible to use topology to deduce the mechanism for recombination” (Goldman and Kauffman 1997, 327).

To apply the fraction of a tangle to molecular biology, the authors make the blanket assumption that all products of recombination, starting from a given unknotted and unlinked form of double-stranded DNA, are closure (numerators) of rational tangles. They assume that the knot or link that are built in the recombination process are obtained by a combination of simple twisting (of the sort that builds new rational tangles from the old) and the addition from single crossings at a smoothing site. The latter operation is what is usually called site-specific recombination by biologists. A crossing is created in place of the smoothing that is the local configuration of the “lined-up” sites. There are two possibilities for such a crossing. In order for the recombination to occur the DNA must twist about to bring these two sites into proximity with the orientations lined up.

Let us now introduce some remarks about the relationship between DNA structure and supercoiling. The DNA may take the form of a ring, and so it can become tangled or knotted. Furthermore, a piece of DNA can break temporarily. While in this broken state, the structure of the DNA may undergo a physical change, and the DNA will finally recombine. Topoisomerase type-I can facilitate the whole process, from the original splicing to the recombination. More generally, DNA topoisomerases play a fundamental role in recombination and genome stability. If it was already recognized at the birth of the double-helix structure of DNA “that unwinding of the intertwined strands would be necessary during semi-conservative replication of the molecule” (see Wang et al. 1990), it is with the discovery of ring-shaped double-stranded DNA that the unwinding problem became a topological one: the two multiply linked parental strands must be unlinked after a round of replication.

Before to go further, we need at this stage to give some fundamental ideas about the Jones polynomial (for a thorough discussion, see Jones 1985; Boi 2021b, c). Discovered by Vaughan Jones in 1985 and denoted by him  $V_L(t)$ , this polynomial is a new knot invariant which proved to be very powerful at differentiating between different equivalence classes of knots, while at the same time being relatively easy to compute. Jones discovered his polynomial while studying von Neumann algebras and gave its interpretation in terms of statistical mechanics (Akutsu and Wadati 1987; Wu 1992). The Jones polynomial  $V_K(t)$  of the knot  $K$  is a Laurent polynomial in  $t$ . More generally, the Jones polynomial can be

<sup>19</sup> Let us give this simple example. In  $\mathbb{R}^3$ , the unknot (the circle  $S^1$ ) is not ambient isotopic to the trefoil knot, since one cannot be deformed into the other through a continuous map of homeomorphisms of the ambient space. Yet, they are ambient-isotopic in  $\mathbb{R}^4$ .

defined for any oriented link  $L$  as a Laurent polynomial in  $t^{1/2}$ , so that the reversing of the orientation of all components of  $L$  leaves  $V_L$  unchanged. In particular,  $V_K$  does not depend on the orientation of the knot  $K$ . For a fixed link, we denote the Jones polynomial simply by  $V$ . There are three standard ways to change a link diagram at a crossing point. The Jones polynomial can be characterized by the following properties:

1. Let  $L$  and  $L'$  be two oriented links which are ambient isotopic, and then,  $V_{L'}(t) = V_L(t)$ .
2. Let  $O$  denote the unknot, then  $V_O(t) = 1$ .
3. The polynomial satisfies the following skein relation  $t^{-1}V_+ - tV_- = (t^{1/2} - t^{-1/2})V_0$ .
4. The Jones polynomial distinguishes between a knot and its mirror image. More precisely, we have the following result. Let  $K_m$  be the mirror image of the knot  $K$ , then  $V_{K_m}(t) = V_K(t^{-1})$ . For example, the Jones polynomial can distinguish the trefoil knot from its mirror image, whereas the Alexander–Conway polynomial cannot.
5. Since the Jones polynomial is not symmetric in  $t$  and  $t^{-1}$ , it follows that in general  $V_{K_m}(t) \neq V_K(t)$ .

It must be stressed that the significance of the Jones polynomial invariant goes far beyond pure mathematics, and in fact, it deeply relates with many topics of microscopic and macroscopic physics as well as with various subjects of the life sciences.

Let us return to the recombination. The process of recombination involves some interesting topological changes in the substrate. It is worth noting that knowledge of the topology of the substrate and product(s) can be used to compute the Jones polynomial of other products (see Murasugi 1996; Kauffman 2001). For instance, a cut in a double-strand DNA, due to a topoisomerase, allows a double-strand DNA to pass through it and recombine. Within the synaptic complex, we can assign local orientation to the respective, small part of the DNA molecule on which the recombinase acts within a circle (Fig. 13).

Suppose we have a single circular DNA molecule that contains a copy of each of the two recombination sites necessary for the reaction. Then, when the enzyme acts on this molecule, the result can be analyzed to determine the effect of the enzyme. We can choose an orientation for the site. When both sites appear on the same circular DNA molecule, these orientations can either point in the same direction, in which case we say that the two have *direct repeats*, or their orientation can point in opposite directions (see Fig. 14); in this case, we have *inverted repeats* (see Fig. 15).

Figure 15 shows the process of recombination with direct and inverted repeats. We have the following steps of the synaptic complex recombination (Fig. 16): (a) The substrate. (b) The pre-recombination synaptic complex (Fig. 16, left); here,  $S$  denotes the substrate tangle, which is unchanged

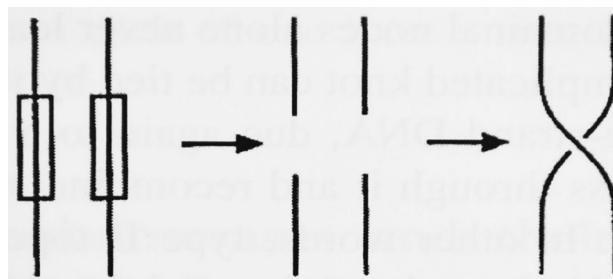


Fig. 13 A possible site-specific recombination

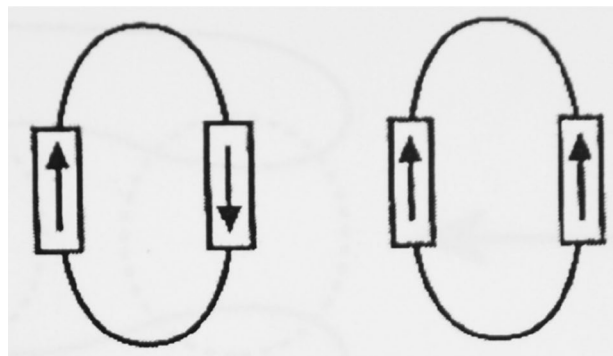


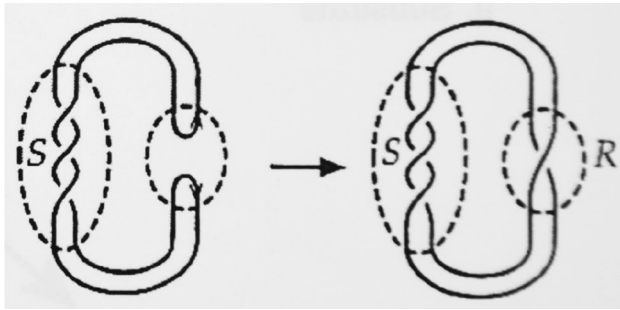
Fig. 14 Two recombination occurring in sites of a circular DNA molecule with different orientations. The type of repeat depends upon the orientation

by the enzyme, and  $T$  stands for the site tangle, where the enzyme acts. (c) The post-recombination synaptic process (Fig. 16, left), thereby the enzyme replaces the site tangle  $T$  with the recombination tangle  $R$ . (d) The product of the recombination, which can be either a knot or a link, according to the above notation, its formula is  $N(T+R)$ , where  $T$  and  $R$  are enzymes determined constants independent of the variable geometry of the substrate  $S$ .

As we just have seen, in the multistep process of recombination of a nicked DNA molecule, the mathematical notion of tangle plays a fundamental role. For the sake of clarity, let us define mathematically the *tangle* (we closely follow Conway 1970; Goldman and Kauffman 1997).

*Description.* On the sphere  $S^2$ , the surface boundary of the three-ball  $B^3$ , take  $2n$  points (see Fig. 17). A  $(n, n)$ -tangle  $T$  is formed by attaching, within  $B^3$ , to these points  $n$  curves, none of which would intersect each other. (The curves should be polygonal.) Suppose that we fix four points on the sphere  $S^2$  (as pictured in Fig. 17)—say, north-east, north-west, south-east, and south-west—to which we attach their coordinates that lie in the  $yz$ -plane. By attaching the end points of two polygonal curves in  $B^3$  to these four points, we can form a tangle. Therefore, if we project this tangle onto the  $yz$ -plane, as in the case of a knot, we have what may be called a *regular diagram of the tangle* (see Fig. 17). The



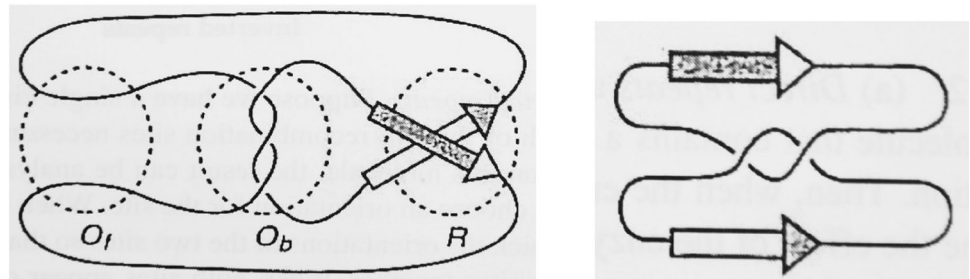
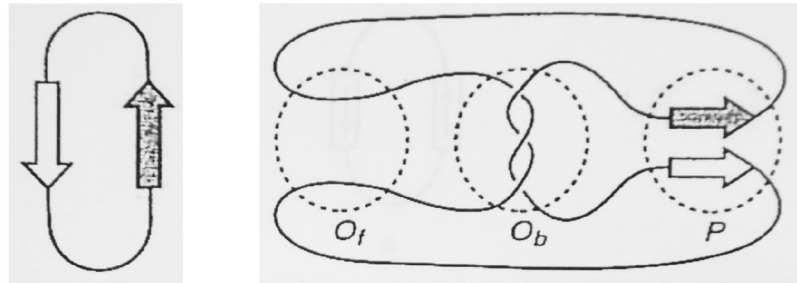


**Fig. 15** (Left) direct repeats. (Right) inverted repeats. **a** Substrate. **b** Pre-recombination synaptic complex. **c** Post-recombination synaptic. **d** Product

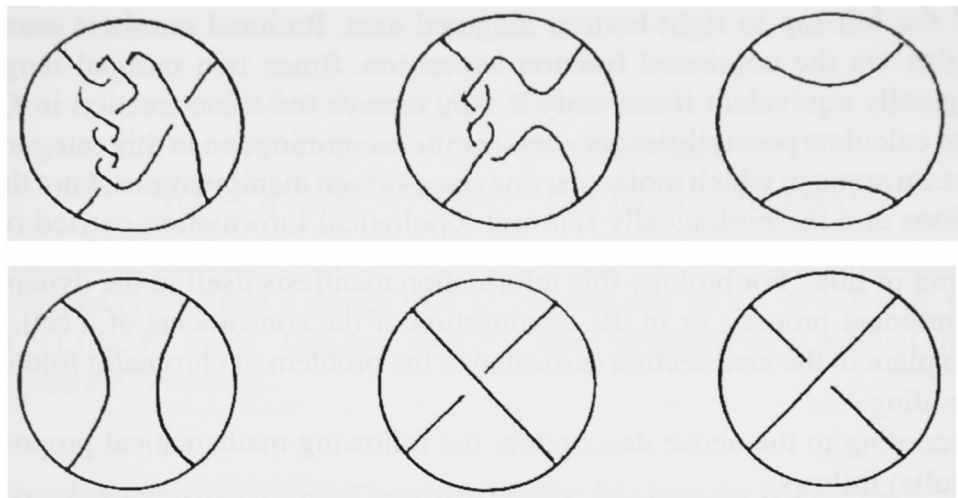
knot (or link) obtained by connecting the points north-west and north-east, south-west and south-east by simple curves outside  $B^3$  is called the *numerator* and is denoted by  $N(T)$ . Similarly, we may connect the points north-west and south-west, north-east and south-east by simple curves outside  $B^3$ , and the subsequent knot (or link) is called the *denominator* and is denoted by  $D(T)$ .

We give some mathematical operations that can be performed on tangle. Let us  $N(Q)$  denote the knot or link obtained by connecting the top two strands of a rationale tangle  $Q$  to each other and the bottom two strands of  $Q$  to each other. Let  $Q + V$  denotes the rationale tangle obtained by adding the two tangles  $Q$  and  $V$  together. In this notation, the

**Fig. 16** Four steps of the synaptic complex recombination with inverted repeats, successively. The recombination process of DNA double-stranded molecule is crucial for the replication of our genetic material and reproduction of cells



**Fig. 17** Representation of tangle diagrams. Two different kinds of rational tangles. Both are twisted  $n$ -tangles (A, B). Four trivial rational tangles (C, D, E, F). The Conway symbol associated with infinity tangle is  $D(0, 0)$



facts that the substrate comes from the tangles  $S$  and  $T$  and the product from the tangles  $T$  and  $R$  can be written in two equations in the three unknowns  $S$ ,  $T$  and  $R$ :  $N(S+T)=\text{substrate}$ , and  $N(T+R)=\text{product}$ . Since we have more variables than equations, we can never determine all three of  $S$ ,  $T$ , and  $R$  from knowing the knotting of the substrate and the product. If we want to know one of the three however, we should be able to determine the other two.

The rational tangles are characterized topologically by values in the extended rational numbers  $\mathbb{Q}^* = \mathbb{Q} \cup \{1/0 = \infty\}$ . An element in  $\mathbb{Q}$  has the form  $\beta/\alpha$  where  $\alpha \in \mathbb{N} \setminus \{0\}$ , ( $\mathbb{N}$  is the natural numbers), and  $\beta \in \mathbb{Z}$  with  $\gcd(\alpha, \beta) = 1$ . Rational tangles themselves are obtained by iterating operations similar to the recombination process itself. The inverse of a tangle is obtained by turning it  $180^\circ$  around the left-top to right-bottom diagonal axis. Rational numbers correspond to tangles via the continued fraction expansion. Since two rational tangles are topologically equivalent if and only if they receive the same fraction in  $\mathbb{Q}^*$ , it is easy to calculate possibilities for site-specific recombination in this category. Here, we have an arena in which molecular enzymes-driven manipulations, knot theoretic operations and the biologically relevant topological information carried out by a knot or link act in a cooperative manner. This brings us directly to the central question of this study: *what is the nature of the topological information carried out by a knot or link?* For biology, this information manifests itself in the dynamics of a recombination process, or in the organization of the constituents of a cell; both are related to the problem of chromatin folding and supercoiling.

According to the previous remarks, the nature of the link between enzymes and topological tangle is encapsulated in the following mathematical propositions:

**Proposition 1** *Almost all the products obtained by the site-specific recombination of trivial knots substrates are rational knots (or links), i.e., two-bridge knots (or links).*

**Proposition 2** *The part of the synaptic complex acted on by an enzyme (recombinase), mathematically within the 3-ball, is a (2, 2)-tangle.*

Therefore, the product is just the replacement of one (2,2)-tangle by another (2,2)-tangle. Thus, for example, a (2,2)-tangle within the circle  $T$  may be replaced by a tangle  $R$  to form a product (Fig. 15). Mathematically, it is perfectly reasonable to consider  $S$  to be a (2,2)-tangle in  $T$ . The numerator of the sum of  $S$  and  $R$  is then the product. Therefore, the following “equation” holds:  $N(S+R)=P$  (the product). Furthermore, we may divide the substrate into the external tangle  $S$  and the internal tangle  $E$ , since the

substrate is the numerator of the sum of  $S$  and  $E$ . Again, we have a quasi-equation holding:  $N(S+E)=S$  (the substrate).

A remarkable fact to be stressed is that tangles depends on the action of enzymes. Thus, an important mathematical assumption, supported by biological observation, is that the tangles  $T$  and  $R$  do not depend on the tangle  $S$ . They only depend on the enzyme that is acting, and not on the knottedness of the molecule it acts on.

There is a very enlightening example to be consider here: the enzyme-topoisomerase *Tn3 resolvase*. We know that it acts on a particular duplex cyclic DNA with direct repeats. Once it has matched up the two sites, it replaces the  $T$  tangle with a single  $R$  tangle and releases the molecule (Fig. 15; see also Fig. 16, bottom left). Once in a while, however, it will repeat the tangle replacement a second time before releasing the molecule. Even more rarely, it can repeat the tangle replacement a number of times, yielding even more complex molecules. From a series of experiments made by biochemists, one can establish what products result when enzymes act, and determine the following equations, where we use the notation for rational tangles

$$N(T+S) = N(1)(\text{the unknot})$$

$$N(T+R) = N(2)(\text{the Hopf link})$$

$$N(T+R+R) = N(2, 1, 1)(\text{the figure – eight knot})$$

$$N(T+R+R+R) = N(1, 1, 1, 1)(\text{the Whitehead link}).$$

From this set of equations, which show how the enzymatic products expressed in terms of operations on tangles generate some types of knots that can be observed experimentally, Sumners (1992) proved that  $S = (-3, 0)$  and  $R = (1)$ . Moreover, he proved that that the expression  $N(S+R+R+R+R) = N(1, 2, 1, 1, 1)$  can ensue (this corresponds to the  $6_2$  knot). This last knot has been observed as a product in many recombination processes.

## Further explanations and interpretations

By the 1980s, it became clear that—although the informational content of the genetic code was embodied in a linear array of bases—it was the three-dimensional structure and the topological condensation in the chromatin-like assembly of the DNA double helix in the chromosomes that ultimately would govern its physiological functions in the cells. This is very likely the crucial point. As an illustration of this point, in perhaps the most striking biological example of ‘forms dictate function’, the two complementary parental strands of DNA must separate during semi-conservative replication to act as the templates for each of the two newly synthesized daughter strands. This discovery leads to the realization that the structure of DNA, while elegant, burdened the cell with

previously unimagined dynamical and topological problems. Although these dynamical and topological problems were originally recognized only for circular molecules, because of the long length of chromosomal DNA, we now know that they apply to linear genomes as well.

The key for finding the solution of these problems seems to lie in the following issues:

1. In the conformational, organizational, and biological roles of the topoisomerases, that, because of their extreme structural and functional complexity, still remains in part to be elucidated.
2. In the DNA supercoiling process, because it links the biological activity of DNA to its tertiary structure and not just its sequence. DNA supercoiling describes a higher order DNA structure. The double-helical structure of DNA entails the interwinding of two complementary strands around one another and around a common helical axis. The writhing of this helical axis in space defines the DNA superhelical structure (DNA tertiary structure). All essential cellular processes seem to be related to the way in which supercoiling is realized.
3. In the three-dimensional organization of the chromatin, which is a nucleoprotein complex and the stuff chromosomes are made of. This organization not only compacts the DNA but also plays a fundamental role in regulating interactions with the DNA during its metabolism.

Condensation of genetic material appears to be a very fundamental mechanism of life. Now, since condensation realizes as a kind of topological embedding of one space, the restrained linear DNA helicoidal-like surface, into another space, the three-dimensional chromosome structure in the cell's nucleus, it seems reasonable to think that topological embeddings and transformations are dynamic processes that are essential for the maintain and the integrity of life (Danchin 1978). One demonstration of that is the fact that the exotic supercoiled forms that double helix can assume are additional complex structures which have an important effect on the molecule's basic (i.e., sequential) structure and its function. For example, supercoiling-induced destabilization of certain DNA sequences can allow the extrusion of cruciform or even the transcriptional activation of eukaryotic promoters. DNA and chromosome organization must fulfill precise topological prerequisite to achieve certain functional processes. In particular, DNA transcription and replication can both be enhanced and regulated by topological supercoiling. It now appears clear, for example, that for replication to be completed, the linking number of the DNA,  $Lk$ , must be reduced from its vast (+) value to exactly zero. In bacteria, DNA gyrase introduces (-) supercoils and thereby removes parental  $Lk$ . Moreover, in

certain cases, the severity of the phenotype can be controlled by changing the level of supercoiling in the cell.

Let us make a few remarks about the general philosophy which underpins this paper. We tried to show the need of working with models that simultaneously integrate geometrical objects, dynamical variables and biological components and their relationships with one another. A multi-level and integrative approach has to essentially take into account the fact that simply knowing the parts list of genes and proteins does not tell us much about how life's many biological processes work. The cellular organization is a complex-dynamic system with hundreds of thousands of bio-molecules interacting with one another to execute life's many functions (Kauffman 1993; Noble 2006). Developments in the mathematical and physical sciences will be very important for addressing complex questions in biology. In the view of these facts, one may foresee that a great deal of the future research on the interface between mathematics, physics, and life sciences will relate to the following two fundamental issues: (1) how did the topology of the double-helix and DNA–protein complexes evolve and (2) why is it so biologically important for the integrity of cells and organisms? These questions arise immediately from the crucial recognition that the topology and dynamics of DNA and macromolecular proteins complexes are essential for the maintenance and integrity of life.

## Conclusion

We have argued that the production of complex living organisms owes much of its working to some topological mechanisms which operate markedly on the three levels of the organization, regulation, and evolution of biological systems. Thus, we can speak of a specific topology of the living acting very dynamically on the substrate space of the physiological and metabolic activities of all complex living organisms (this idea was originally stated by Waddington (1968) and Thom (1972, 1989), and thereafter, in more philosophical terms, by Simondon (2005); see also Rosen (1970) and Goodwin and Webster (1996)). There are geometrical (local) transformations and topological (global) remodeling which seems to play a central role in the enhancement and modulation of the required spatial changes occurring in the organism during its embryogenetic development and the cell differentiation (leading to the formation of tissues and organs). There are also, upstream, some geometric transformations and topological remodeling of nuclear structures that control and orchestrate the conditions of expressivity of genes and contemporarily the systems of epigenetic regulation at the level of the assembling of chromatin and that of the organization of chromosomes (Kimmins and Sassoni-Corsi 2005, Ridgway and Almouzni 2001).

In this paper, we tried to show that the conformational plasticity of biological systems, at the genome and epigenome levels, mainly depends on the topological action by specific enzymes, which effectively can link structures to dynamics and changes of forms to the emergence of new functions. In our view, the employment of differential geometry and topological knot theory does not restrict to model the *in vitro* observed properties and the artificially supposed mechanisms of molecular structures and functions. What is required is much more the understanding of how some precise mathematical operations and physical processes participate in and in certain case promote the formation and evolution of specific biological structures and functions. The example we studied here of the link between the topological knot theory and the folding of the three-dimensional structures of protein-DNA complexes clearly illustrate a deep and active connection interaction between topology, physics, and biology.

In this study, we set the emphasis on the following four most relevant work assumptions: (1) that topological changes and dynamical processes provide a nexus for mathematics and biology. (2) That these changes and processes occur in the framework of different fluctuations and instabilities affecting some physical parameters like temperature, energy and possibly other thermodynamically varying conditions (Nicolas and Prigogine 1977), and in diverse case, the topological objects and operations assure a certain structural and functional stabilization; actually, this hypothesis was leaved implicit, because it needs to be investigated and clarified much further. (3) That certain geometric properties and topological patterns are essential for the organization and growth of biological systems. In order, these properties and patterns can produce real biological activities, it is required that they must be effectively combined with specific physical processes occurring in the organism, conceived as an open complex system and an autonomous self-organizing system at once. (4) That those properties and patterns provide the organism with adaptative plasticity and robust functionality at micro, meso, and macro scales.

Thus, we can tentatively claim that the topological mechanisms discussed here operate on the organization, regulation, and evolution of biological systems, primarily at the molecular and macromolecular level, but also that geometrical modifications (bending, writhing, and twisting) and topological remodeling (coiling, knotting, and untying) apparently play a central role during embryogenetic development and cell differentiation (Furlan-Margaril and Recillas-Targa 2011).

From a more theoretical point of view, it is clear that the genetic causality theory has several limitations, both intrinsic because of the multi-level complexity of biological processes and extrinsic in that it disregards the influence of the phenotype on the genotype and in particular the possibility

that certain acquired characteristics can be inherited. In a sense, we can say that the molecular biological conception of recent decades has limited or even misleadingly impacted our vision of the living world. New ideas are needed if we are to succeed in unraveling multifactorial genetic, epigenetic, and environmental causation at higher levels of physiological function and so to explain fundamental living phenomena that genetics alone is unable to explain (see Noble 2006; Boi 2017). Even from the study of the nuclear genome activity and the related cell functions, which is the one we principally have addressed in this paper, it is possible to conclude that (1) structural plasticity and biological functionality are deeply related and multi-level (the chromatin remodeling and functionality is a clear illustration of this fact (see Felsenfeld and Groudine 2003)), (2) the biological information is inherently spatial and temporal (think for instance of the proteins activity whose biological functions are sensitive to their topological folding in the cell space), it is not unidirectional, and it essentially evolve following a complex and changing network-like organization, (3) the theory of inheritance need a deep conceptual reformulation (see Holliday 1987; Danchin and Charmantier 2011)), first because it can no more rest on the believe that DNA is the sole carrier of inheritance, and second because what is transmitted is not only the replicated part of the genetic material but also other relevant parts and properties of the cellular and organismic metabolism (see Dyson 1985; Misteli 2007), and (4) gene ontology is lacking and confusing without considering other fundamental levels of the organization and regulation of the living systems (see McClintock 1984; Jaenisch and Bird 2003; Cavalli and Heard 2019).

**Acknowledgements** We wish to thank the referees for very useful comments which allowed the revision of several inaccuracies concerning some mathematical and biological statements and hence the improvement of this article. We also would like to thank Andras Paldi, Moncef Ladjimi, Hans Liljenström, Jürgen Jost, and Carlos Lobo for helpful discussions.

## References

- Adams CC (2000) The knots book: an elementary introduction to the mathematical theory of knots. W. H. Freeman, New York
- Akutsu Y, Wadati M (1987) Knot invariants and critical statistical systems. *J Phys Soc Jpn* 56:839–842
- Alberts B (2003) DNA replication and recombination. *Nature* 421:431–435
- Andersen JE, Penner RC, Reidys CM, Waterman MS (2013) Topologically classification and enumeration of RNA structures by genus. *J Math Bio* 67(5):1261–1278
- Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181(4096):223–230
- Bates A, Maxwell A (2005) DNA topology, 2nd edn. Oxford University Press, Oxford

- Begun A, Liubimov S, Molochkov A, Niemi AJ (2021) On topology and knotted entanglements in protein folding. *PLoS ONE* 16(1):1–17
- Bian Q, Belmont AS (2012) Revisiting higher order large-scale chromatin organization. *Curr Opin Cell Biol* 24(3):359–366
- Birman JS (1974) Braids, links, and mapping class groups. Princeton University Press, Princeton
- Boi L (2005) Topological knots models in physics and biology. In: Boi L (ed) Geometries of nature, living systems and human cognition. New interactions of mathematics with natural sciences and humanities. World Scientific, Singapore, pp 203–278
- Boi L (2006) Mathematical knot theory. In: Françoise J-P, Naber G, Sun TS (eds) Encyclopedia of mathematical physics, vol 3. Elsevier, Oxford, pp 399–406
- Boi L (2007a) Geometrical topological modeling of supercoiling in supramolecular structures. *Biophys Rev Lett* 2(3):1–13
- Boi L (2007b) Modelling supercoiling in biological structures. In: Di Gesù V, Lo Bosco G, Maccarone MC (eds) Modelling and simulation in science. World Scientific, Singapore, pp 187–200
- Boi L (2007c) Sur quelques propriétés géométriques globales des systèmes vivants. *Bull D'histoire D'Épistémol Sci Vie* 14:71–113
- Boi L (2009) Epigenetic phenomena, chromatin dynamics, and gene expression. *New theoretical approaches in the study of living systems. Biol Forum* 101(3):405–442
- Boi L (2011a) When topology meets biology ‘for life’. Remarks on the way in which topological form modulates biological function. In: New trends in geometry and its role in the natural and life sciences. Imperial College Press, London, pp 241–303
- Boi L (2011b) Plasticity and complexity in biology: topological organization, regulatory protein networks and mechanism of gene expression. In: Terzis G, Arp R (eds) Information and living systems. Philosophical and Scientific Perspectives. The MIT Press, Cambridge, pp 205–250
- Boi L (2017) The interlacing of upward and downward causation in complex living systems: on interactions, self-organization, emergence, and wholeness. In: Paolini Paoletti M, Orilia F (eds) Philosophical and scientific perspectives on Downward causation. Routledge, London, pp 180–203
- Boi L (2021a) Geometrical modeling of DNA and the structural complexity of the chromosome. *J Biophys* (forthcoming)
- Boi L (2021b) A topological and dynamical approach to the study of complex living systems. In: Albeverio S, Mastrogiacomo E (eds) Complexity and emergence. Springer, Heidelberg, pp 57–104
- Boi L (2021c) Knots, diagrams, and kid’s shoelaces: on spaces and their forms. In: Boi L, Lobo C (eds) When form becomes substance. Power of gesture, diagrammatical intuition and phenomenology of space. Birkhäuser, Basel, pp 137–208
- Boles CT, White JH, Cozzarelli NR (1990) Structure of plectonemically supercoiled DNA. *J Mol Biol* 213(4):931–951
- Bon M, Vernizzi G, Orland H, Zee A (2008) Topological classification of RNA structures. *J Mol Biol* 379:900–911
- Brunello L, Levens D, Gupta A, Kouzine F (2012) The importance of being supercoiled: How DNA mechanic regulate dynamic processes. *Biophys Acta (BBA) Gene Regul Mech* 1819(7):632–638
- Buck D (2009) DNA topology. *Proc Symp Appl Math* 66:1–33
- Buck D, Valencia D (2011) Characterization of knots and links arising from site-specific recombination of twist knots. *J Phys A* 44(4):1–36
- Burde G, Zieschang H (2003) Knots, 2nd edn. de Gruyter, Berlin
- Carbone A, Gromov M (2001) Mathematical slices of molecular biology. *Gazette des Mathématiciens. Soc Math France* 8:11–80
- Cavalli G, Heard E (2019) Advances in epigenetics link genetics to environment and disease. *Nature* 571:39–68
- Conway JH (1970) An enumeration of knots and links, and some of their algebraic properties J. In: Leech (ed) Computational problems in abstract algebra. Pergamon Press, Oxford, pp 329–358
- Cozzarelli NR, Spengler SJ, Stasiak A (1985) The stereostructure of knots and catenanes produced by phase  $\lambda$  integrative recombination: implications for mechanism and DNA structure. *Cell* 42:325–334
- Cozzarelli NR (1992) Evolution of DNA topology: implications for its biological role. In: New scientific applications of geometry and topology, PSAM, vol 45, Amer. Math. Soc
- Cremer T et al (2004) Higher order chromatin architecture in the cell nucleus: on the way from structure to function. *Biol Cell* 96:555–567
- Culler M, Gordon MCA, Leucke J, Shalen PB (1987) Dehn surgery on knots. *Ann Math* 125(2):237–300
- Danchin A (1978) *Ordre et dynamique du vivant*. Éditions du Seuil, Paris
- Danchin E, Charmantier A (2011) Beyond DNA: Integrating inclusive inheritance into an extended theory of evolution. *Nat Rev Gen* 12:475–486
- Darcy IK, Levene SD, Scharein RG (2014) Introduction to DNA topology. In: Jonoska N, Saito M (eds) Discrete and topological models in molecular biology. Springer, Heidelberg, pp 327–345
- Dehn M (1910) Über die topologie des dreidimensionalen raumes. *Math Ann* 69(1):137–168
- Dixon JR, Gorkin DV, Ren B (2016) Chromatin dynamics: the unit of chromosome organization. *Mol Cell* 62(5):668–680
- Douglas J (1931) Solution of the problem of Plateau. *Trans Am Math Soc* 33(1):263–321
- Durickovic B, Goriely A, Maddocks JH (2013) Twist and stretch of helices explained via the Kirchhoff-Love rod model of elastic filaments. *Phys Rev Lett* 111:108103–108105
- Dyson F (1985) *Origins of life*. Cambridge University Press, Cambridge
- Elhamdadi M, Hajjij M, Istvan K (2020) Framed knots. *Math Intell* 42:7–22
- Ernst C, Sumners DW (1990) A calculus for rational tangles: applications to DNA recombination. *Math Proc Camb Math Soc* 108(3):489–515
- Felsenfeld G, Groudine M (2003) Controlling the double helix. *Nature* 421:448–453
- Flapan E, Grevet J, Li Q, Sun CD, Wong H (2014) Knotted and linked products of recombination on  $T(2, n)\#T(2, m)$  substrates. *J Korean Math Soc* 51(4):817–836
- Flapan E, He A, Wong A (2019) Topological description of protein folding. *Proc Natl Acad Sci USA* 116(19):9360–9369
- Forterre P, Gribaldo S, Gabelle D, Serre M-C (2007) Origins and evolution of DNA topoisomerases. *Biochimie* 89(4):427–446
- Fuller FB (1978) Decomposition of the linking number of a closed ribbon: a problem from molecular biology. *Proc Natl Acad Sci USA* 75(8):3557–3561
- Furlan-Margaril M, Recillas-Targa F (2011) Chromatin remodeling and epigenetic regulation during development. In: Chimal-Monroy J (ed) Topics in animals and plant development: from cell differentiation to morphogenesis, pp 221–247
- Goldman JR, Kauffman LH (1997) Rational tangles. *Adv Appl Math* 18(3):300–332
- Goodwin B, Webster G (1996) *Form and transformation: generative and relational principles in biology*. Cambridge University Press, Cambridge
- Gordon CM (2006) Some aspects of classical knot theory. In: Hausmann JC (ed) *Knot theory, lecture notes in mathematics*, vol 685. Springer, Heidelberg, Berlin, pp 1–60
- Gromov M (2011) Crystals, proteins, stability and isoperimetry. *Bull Am Math Soc (NS)* 48(2):229–257
- Hinde E, Cardarelli F, Digman MA, Gratton E (2012) Changes in chromatin compaction during the cell cycles revealed by

- micrometer-scale measurement of molecular flow in the nucleus. *Biophys J* 102(3):691–697
- Hirano T (2016) Condensin-based chromosome organization from bacteria to vertebrates. *Cell* 164(5):847–857
- Holliday R (1987) The inheritance of epigenetic defects. *Science* 238:163–170
- Huang FW, Reidys CM (2015) Shapes of topological RNA structures. *Math Biosci* 270:57–65
- Huang FW, Reidys CM (2016) Topological language for RNA. *Math Biosci* 282:109–120
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33:245–254
- Jones VFR (1985) A polynomial invariant for knots via von Neumann algebras. *Bull Am Math Soc* 12:103–111
- Jost J (2019) *Biologie und mathematik*. Springer, Berlin, Heidelberg
- Jost D, Carrivain P, Cavalli G, Vaillant C (2014) Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res* 42(15):9541–9549
- Kauffman LH (1987) *On knots*. Princeton University Press, Princeton
- Kauffman LH (1990) An invariant of regular isotopy. *Trans Am Math Soc* 318(2):417–471
- Kauffman S (1993) *The origins of order: self-organization and selection in evolution*. Oxford University Press, Oxford
- Kauffman LH (2001) *Knots and physics*, world scientific, series on knots and everything, vol 1. World Scientific, London
- Kauffman LH (2005) *Knots*. In: Boi L (ed) *Geometries of nature, living systems and human cognition. The new interactions of mathematics with natural sciences and the humanities*. World Scientific, Singapore, pp 131–202
- Kauffman LH, Lambropoulou S (2004) On the classification of rational tangles. *Adv Appl Math* 33(2):199–237
- Képès F, Vaillant C (2003) Transcriptional-based solenoidal model of chromosomes. *Complexus* 1(4):171–180
- Kervaire M (1965) Les nœuds de dimensions supérieures. *Bull Soc Math France* 93:225–271
- Kimmins S, Sassoni-Corsi P (2005) Chromatin remodeling and epigenetic features of germ cells. *Nature* 434:583–589
- Kirby R (1978) A calculus for framed links in  $S^3$ . *Invent Math* 45(1):35–56
- Kitano H (2004) Biological robustness. *Nat Rev Genet* 5(11):826–837
- Lal A et al (2016) Genome scale patterns of supercoiling in a bacterial chromosome. *Nat Commun* 7(1):11055–11163
- Lickorish WBR (1997) *An introduction to knot theory, graduate texts in mathematics*, vol 175. Springer, Heidelberg
- Lodish H, Berk A, Zipursky A et al (2000) *Molecular cell biology*, 4th edn. W. H. Freeman, New York
- Mazur B (2004) Perturbations, deformations, and variations (and “near-misses”) in geometry, physics, and number theory. *Bull Am Math Soc (NS)* 41(3):307–336
- McClintock M (1984) The significance and responses of the genome to challenge. *Science* 226:792–801
- McGinty RK, Tan S (2015) Nucleosome structure and function. *Chem Rev* 115:2255–2273
- Misteli T (2007) Beyond the sequence. Cellular organization of genome function. *Cell* 128(4):787–800
- Murasugi K (1996) *Knot theory and its applications*. Birkhäuser, Boston
- Muskhelishvili G, Travers A (2016) The regulatory role of DNA supercoiling in nucleoprotein complex assembly and genetic activity. *Biophys Rev* 8(Suppl. 1):5–22
- Nicolas G, Prigogine I (1977) *Self-organization in nonequilibrium systems: from dissipative structures to order through fluctuations*. Wiley, New York
- Noble D (2006) *The music of life. Biology beyond the genome*. Oxford University Press, Oxford
- Noble D (2008) Genes and causation. *Phil Trans R Soc Lond A* 366(1878):3001–3015
- Ochs F et al (2019) Stabilization of chromatin topology safeguards genome integrity. *Nature* 574:571–574
- Ophl WF, Roberts GW (1978) Topological considerations in the theory of replication of DNA. *J Math Biol* 6:383–402
- Penner RC (2016) Moduli spaces and macromolecules. *Bull Am Math Soc* 53:217–269
- Penner RC, Waterman MS (1993) Spaces of RNA secondary structures. *Adv Math* 101(1):31–49
- Peselis A, Serganov A (2014) Structure and function of pseudoknots involved in gene expression control. *Wiley Interdisc Rev RNA* 5(6):803–822
- Porter LL, Looger LL (2018) Extant fold-switching proteins are widespread. *Proc Natl Acad Sci USA* 115(23):5968–5973
- Ramam V, Shendure J, Duan Z (2016) Understanding Spatial Genome Organization: Methods and Insights. *Genom Proteom Bioinform* 14(1):7–20
- Reidemeister K (1927) *Elementare begründung der knotentheorie*. Abh Math Sem Univ Hamburg 5(1):2432
- Reidemeister K (1932) *Knotentheorie*. Springer, Heidelberg/Berlin/New York
- Ricca RL, Nipoti B (2011) Gauss’s linking number revisited. *J Knot Theory Ramific* 20(10):1325–1343
- Ridgway P, Almouzni G (2001) Chromatin assembly and organization. *J Cell Sci* 114:2711–2722
- Roca J (1998) Topoisomerases. *Adv Genome Biol* 5:463–485
- Rolfen D (1976) *Knots and links, mathematical lecture series, vol 7*. Publish or Perish, Huston
- Rosen R (1970) *Dynamical systems theory in biology*. Wiley, New York
- Scherrer K, Jost J (2007) Gene and genon concept: coding versus regulation. A conceptual and information-theoretic analysis of genetic storage and expression in the light of modern molecular biology. *Theory Biosci* 126(2):65–113
- Seifert H (1935) Über die Geschlecht von Knoten. *Math Ann* 110(1):571–592
- Sergei MM (2001) *DNA topology: fundamentals, encyclopedia of life sciences*. Nature Publishing Group, Berlin, pp 1–11
- Simondon G (2005) *L’individuation à la lumière des notions de forme et d’information*, Jérôme Million, Paris
- Spera M (2006) A survey on the differential and symplectic geometry of linking numbers. *Milan J Math* 74:139–197
- Strick TR, Allemand J-F, Bensimon D, Croquette V (1998) Behavior of Supercoiled DNA. *Biophys J* 74:2016–2028
- Summers DW (1990) Untangling DNA. *Math Intell* 12(3):71–80
- Summers DW (1992) *Knot theory and DNA*. In: *New scientific applications of geometry and topology, PSAM, 45*, Amer Math Soc, pp 39–72
- Sutormin DA et al (2021) Diversity and Functions of Type II Topoisomerases. *Acta Natur* 13(1):59–75
- Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev* 9:465–476
- Theimer CA, Blois CA, Feigon J (2005) Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* 17(5):671–682
- Thom R (1972) *Stabilité structurelle et morphogénèse*. Benjamin, New York
- Thom R (1989) *Modèles mathématiques de la morphogénèse*. Christian Bourgeois, Paris
- Vazques M, Summers DW (2004) Tangles analysis of Gin site-specific recombination. *Math Proc Camb Phil Soc* 136(565):565–582
- Venkata RY, Bansal M (2017) DNA structural features of eukaryotic TATA-containing and TATA-less promoters. *FEBS Open Bio* 7(3):324–334

- Villota-Salazar NA, Mendoza-Mendoza A, Gonzáles-Prieto JM (2016) Epigenetics: from the past to the present. *Front Life Sci* 9(4):347–370
- Vologodskii AV (1992) *The topology and physics of circular DNA*. CRC Press, Boca Raton, FL
- Waddington CH (1957) *The strategy of the genes*. Routledge, London
- Waddington CH (ed) (1968) *Toward a theoretical biology*. Routledge, London, pp 1968–1969
- Wang JC (1996) DNA topoisomerases. *Ann Rev Biochem* 65:635–692
- Wang JC, Caron PR, Kim RA (1990) The role of DNA topoisomerases in recombination and genome stability: a double-edged sword. *Cell* 62:403–406
- White JH (1989) *An introduction to the geometry and topology of DNA structures*. CRC Press, Boca Raton
- White JH, Cozzarelli NR, Bauer WR (1988) Helical repeat and linking number of surface-wrapped DNA. *Science* 241:323–327
- Wu FY (1992) Knot theory and statistical mechanics. *Rev Mod Phys* 64(4):1099–1129
- Zeeman EC (1960) Unknotting spheres. *Ann Math* 72:350–361
- Zeeman EC (1965) Twisting spun knots. *Trans Am Math Soc* 115:471–495
- Zhurkin VB, Norouzi D (2021) Topological polymorphism of nucleosome and folding of chromatin. *Biophys J* 120(4):577–585

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.