

UNIVERSITA' DEGLI STUDI DI NAPOLI FEDERICO II



Department of Biology

Ph.D. in Biology

XXIX Cycle

Impact of DNA Topology on cellular metabolism: *in vivo* and *in vitro* approaches to study chromatin dynamics and DNA Topoisomerases.

Coordinator:

Chiar.mo Prof. Salvatore Cozzolino

Tutor:
Dr Anna Valenti

Ph.D. Student:
Dr Valeria Visone

INDEX

Summary	3
Sommario	4
1.Introduction	6
• 1.1 DNA Topology and Transcription	6
• 1.2 Structural diversity of DNA	7
• 1.3 Understanding Reverse Gyrase function	10
• 1.4 DNA Topoisomerases and their poisoning	11
References	14
2.Papers	20
• 2.1 Massively Systematic Transcript End Readout (MASTER): Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields	20
• 2.2 Comparison of Plasmid DNA Topology at nanoscale level	43
• 2.3 Heterologous expression of thermophilic reverse gyrase in an archaeal model system by exploiting a new thermostable protein tag	60
• 2.4 A new Daunomycin-Oligoarginine conjugate: biochemical characterization	70
• 2.5 Chromatin Structure and Dynamics in Hot Environments: Architectural Proteins and DNA Topoisomerases of Thermophilic Archaea	85
3.Conclusions and Future Perspectives	118
• 3.1 DNA Topology and Transcription	118
• 3.2 Structural diversity of DNA	119
• 3.3 Understanding Reverse Gyrase function	121
• 3.4 DNA Topoisomerases and their poisoning	123
References	125

Summary

DNA topology describes the tertiary conformations of DNA and it influences all the fundamental biological processes such as transcription, replication and recombination. DNA topoisomerases are essential enzymes able to covalently modify DNA topology and they are emerging as important factors in a wide range of fundamental metabolic processes in both the nuclear and mitochondrial genomes. Furthermore the study of topoisomerases is therapeutically relevant to cancer, immune disorders and neurological diseases. Topoisomerases introduce transient DNA breaks using a trans-esterification mechanism, which is highly reversible and minimizes the risks to genome stability that would otherwise occur owing to strand breakage. Furthermore the topological state of DNA is not equal among different classes of organisms. In mesophilic organisms DNA is negatively supercoiled while positive supercoiling is a peculiar mark of thermophilic organisms living at high temperatures.

During my PhD I analyzed different aspects related to DNA topology, in particular:

- the relationship between DNA topology and transcription was investigated. DNA supercoiling has been shown to be essential for transcriptional regulation. My analysis demonstrates that DNA topology is a determinant of the selection of transcription starting site.
- the structural diversity of negatively and positively supercoiled DNA molecules was characterized by biochemical assays and atomic force microscopy. My data demonstrate that supercoiling strongly influences DNA structure. In particular the positive supercoiling is more stable than negative protecting DNA from nuclease degradation and helix breathing. These data support the potential role of positive supercoiling in DNA stability maintenance.
- the biochemical properties of topoisomerases inhibitor Daunomycin and its derivative were investigated. It's known that human topoisomerases are important target of several antitumor agents such as anthracyclines. However their use in chemotherapy is limited by a lot of side effects especially cardiotoxicity. One approach to overcome this problem consists in testing new molecular transporters (such as oligopeptides) that do not alter the characteristics of the drugs but reduce the side effects. My results demonstrate that Daunomycin conjugated with an oligopeptide containing six arginines retains the same biochemical properties of the free drug, representing a good alternative to Daunomycin.
- I described the *in vivo* detection of reverse gyrase by using a thermostable protein-tag in the thermophilic organism *Sulfolobus islandicus*. Reverse gyrase is a DNA topoisomerase with the unique capability to introduce positive supercoils in DNA. Although a lot of *in vitro* studies have been performed to understand the mechanical properties of positive supercoiling reaction, its biological role *in vivo* remains unclear. This new thermostable protein-tag will be useful for reverse gyrase analysis both *in vitro* and *in vivo*.

Sommario

La topologia del DNA descrive le conformazioni terziarie del DNA ed influenza tutti i principali processi biologici quali trascrizione, replicazione e ricombinazione. Le DNA topoisomerasi sono enzimi essenziali capaci di modificare covalentemente la topologia del DNA e stanno emergendo come importanti fattori in un'ampia gamma di processi metabolici fondamentali sia nei genomi nucleari che in quelli mitocondriali. Inoltre, lo studio delle topoisomerasi è rilevante nella ricerca terapeutica su cancro, disturbi del sistema immunitario e malattie neurologiche. Le topoisomerasi introducono rotture transitorie del DNA utilizzando un meccanismo di transesterificazione, che è altamente reversibile e minimizza i rischi nei riguardi della stabilità genomica che potrebbero verificarsi a causa della rottura del filamento. Lo stato topologico del DNA non è uguale tra le diverse classi di organismi. Negli organismi mesofili il DNA è superavvolto negativamente mentre il superavvolgimento positivo è una caratteristica peculiare degli organismi termofili che vivono ad alte temperature.

Durante il mio Dottorato di ricerca ho analizzato diversi aspetti legati alla topologia del DNA, in particolare:

- Ho analizzato la relazione tra topologia e trascrizione del DNA. Il superavvolgimento del DNA ha dimostrato di essere essenziale per la regolazione trascrizionale. L'analisi condotta dimostra che la topologia del DNA è un determinante della selezione del sito di inizio della trascrizione.
- La diversità strutturale di molecole di DNA negativamente e positivamente superavvolte è stata analizzata attraverso saggi biochimici e microscopia a forza atomica. Nel loro insieme, i miei dati dimostrano che il superavvolgimento influenza fortemente la struttura del DNA. In particolare il superavvolgimento positivo rende più stabile il DNA proteggendolo dalla degradazione da parte di nucleasi nonché dal “breathing” della doppia elica. Questi dati supportano il ruolo potenziale del superavvolgimento positivo nel mantenimento della stabilità del DNA, in particolare alle alte temperature.
- Ho analizzato le proprietà biochimiche dell'inibitore di topoisomerasi Daunomicina e di un suo derivato. E' noto che le topoisomerasi umane sono importanti bersagli di diversi composti antitumorali come le antracicline. Tuttavia il loro uso in chemioterapia è limitato da numerosi effetti collaterali ed in particolare la cardiotoxicità. Uno degli approcci utilizzati per superare questo problema consiste nel testare nuovi trasportatori molecolari (come oligo-peptidi) che non alterino le caratteristiche delle droghe ma ne riducano gli effetti collaterali. I miei risultati dimostrano che la Daunomicina derivatizzata con un oligo-peptide contenente sei arginine conserva le proprietà biochimiche della droga da sola, andando a rappresentare una buona alternativa alla Daunomicina.
- Ho analizzato la marcatura *in vivo* della girasi inversa attraverso l'utilizzo di un “tag” termostabile nell'archaeon termofilo *Sulfolobus islandicus*. La girasi inversa è una DNA topoisomerasi con la capacità unica di introdurre superavvolgimenti positivi nel DNA. Sebbene

siano stati realizzati molti studi *in vitro* al fine di comprendere le proprietà meccaniche della reazione di superavvolgimento positivo, il ruolo biologico *in vivo* della girasi inversa resta ancora da chiarire. Questo nuovo tag termostabile sarà utile per nuove analisi della girasi inversa sia *in vitro* che *in vivo*.

1. Introduction

1.1 DNA Topology and Transcription

Typical double-stranded DNA consists of a helix with a pitch of one turn per 10.5 base pairs. Both underwinding and overwinding of the DNA double helix induce twisting and coiling of the helix unless the DNA strand can rotate freely. The coils thus formed are termed negative and positive supercoils, respectively. Current views of the mechanisms underlying transcriptional regulation rely on a concept originally proposed by Jacob and Monod: regulation through cis-elements on DNA and trans-acting factors that bind to the elements (Jacob and Monod 1961). However, DNA supercoiling can modulate accessibility of trans-acting factors to cis-elements. Transcription is an asymmetric process: only one strand of the DNA double helix is copied into RNA. To achieve this, the double helix must be locally unwound. Moreover, DNA supercoiling can affect transcription in chromatin context in eukaryotes.

Topological aspects of DNA structure arise primarily from the fact that the two DNA strands are repeatedly intertwined. Untangling these two strands, which occurs in all major genetic processes may prove rather difficult. In the simplest case of a linear DNA in solution, untangling is possible due to the free rotation of the ends of the DNA. However, for all natural DNAs, free end rotation is either restricted or forbidden altogether. Consequently, untangling the two DNA strands becomes topologically impossible.

In the last few years, new approaches have reinforced the strong relationship between DNA topology and transcription and how DNA topology provides an additional level of transcriptional regulation and gene expression. DNA supercoiling has been shown to be essential for transcriptional regulation in bacteria (Sobetzko 2016) since more than half of the genes is sensitive to DNA supercoiling. DNA supercoiling can affect transcription in multiple ways: it supports the local untwisting of DNA facilitating the transcription initiation by RNA polymerase (RNAP); it can affect the local DNA geometry and thus modulate transcription factor binding.

For example, it is well known that in *E. coli* DNA is negatively supercoiled and DNA supercoiling levels are under a tight homeostatic control. Free DNA superhelicity is furthermore buffered by abundant nucleoid-associated-proteins (NAPs) that wrap DNA and therewith constrain DNA supercoils. Although topoisomerases and NAPs balance the levels of global DNA supercoiling, the overall level of negative superhelicity continuously decreases from exponential to stationary growth phase. However, the mechanism creating thousands of diverse temporal gene expression patterns coordinated by DNA supercoiling remains unclear.

Sobetzko and colleagues showed that a specific chromosomal arrangement of genes is able to modulate the local levels of DNA supercoiling at gene promoters via transcription-coupled DNA supercoiling (TCDS) in *E. coli*. These findings also provide a consistent explanation for the strong positive coupling of temporal gene expression patterns of neighboring genes. Furthermore, using comparative genomics,

this study provides evidence that TCDS is a driving force for the evolution of chromosomal gene arrangement patterns in other *Enterobacteriaceae* (Corless and Gilbert 2016). Mounting evidence supports a central role for transcription dependent DNA supercoiling in disrupting chromatin structure at all scales. The influence of DNA supercoiling on the expression of neighboring genes has also been inferred at the kilobase scale by linking co-transcriptional regulation to gene orientation (Meyer and Beslon 2014). In this study divergent promoters show mutual elevation of expression, as these promoters drive under-wound DNA into their neighbor, whereas convergent promoters show mutual repression, which may be due to the presence of over-wound DNA.

A lot of other studies demonstrate that DNA supercoiling functions as a regulator of prokaryotic transcription (Pruss and Drlica 1989). First, supercoiling affects transcription *in vitro*. Some promoters have an optimum level of supercoiling for their transcription and the level is different for different promoters. Second, supercoiling plays a regulatory role *in vivo*. Genetic studies have shown that mutations in the topoisomerases responsible for the homeostatic control of DNA supercoiling affect transcription *in vivo* (Margolin et al. 1985; Rudd et al. 1987). Two rate-limiting steps are known for prokaryotic transcription. One is formation of an RNA polymerase-DNA open complex and the other is promoter clearance. Because negative supercoiling favors the unwinding of the DNA double helix that is required for formation of the open complex, it is expected to increase the rate of transcription for promoters in which open complex formation is rate limiting. Indeed most genes are activated by increased negative supercoiling. However, transcription of *gyrA* and *gyrB* encoding the subunits of DNA gyrase is induced by relaxation of DNA. It has been proposed that promoter clearance but not open complex formation is the rate-limiting step for these promoters (Menzel and Gellert 1987). The superhelical state of DNA is known to change depending on the growth conditions of cells (Drlica 1992). For example, nutrient downshift and stationary growth phase cause a decrease in the extent of negative supercoiling, while high osmolarity leads to an increase in the negative supercoiling of DNA. DNA supercoiling also changes transiently during heat shock. The heat stress induces rapid relaxation of negative supercoils and then DNA topology returns back to the original state with negative supercoiling. In response to the relaxation, most genes are repressed but some specific or stress genes are induced (Hirose and Kuniharu 2005).

1.2 Structural diversity of DNA

Bacterial DNA is folded into a compact nucleoid body, which is a dynamic entity that alters its overall structure in response to changes in the bacterial growth rate and growth phase. The nucleoid DNA is thought to be packed in numerous loops, which are compacted at least 1000-fold to fit into the bacterial cell. Yet, this tightly folded organization must not impair its function. The ordered DNA packing, as well as the dynamics in the evolution of nucleoid structure, are achieved through different mechanisms and

are correlated with changes in the distribution and utilization of DNA supercoiling (Japaridze et al. 2017). DNA supercoiling is thought to play key roles in genetic processes in the cell (Sinden 1994). One fundamental feature of DNA supercoiling is that distantly separated DNA regions can be brought in close proximity. Bringing two distant sites into proximity is required for DNA recombination (Craigie 1986; Benjamin and Cozzarelli 1986; Echols 1986; Summers 1995), for the control of DNA supercoiling by DNA topoisomerases (Wang 1985; Gellert 1987; Zechiedrich and Osheroff 1990), and for gene regulation through DNA looping (Adhya 1989; Hochschild 1990; Schleif 1992).

Furthermore, the chromatin structure is regulated by a complex interplay of abundant nucleoid-associated proteins (NAPs), DNA translocases, topoisomerases, and condensing agents. It is well known that, during the bacterial growth cycle, both the relative and absolute concentrations of NAPs change substantially and coordinately with changing global supercoiling levels (Sobetzko 2016). The link between DNA supercoiling and genetic regulation, and the role of global supercoiling in chromatin organization is under intense examination.

DNA supercoiling has been analysed by a variety of approaches and only the most common are described in this paragraph. One of the first methods used historically was been the titration of supercoiling density by sedimentation in ethidium bromide sucrose density gradients. This method is grounded in two facts:

- (1) because supercoiled molecules are more compact than relaxed ones, they sediment faster through a sucrose density gradient;
- (2) ethidium bromide intercalates between the stacked DNA base pairs, unwinding the double helix by 26° per intercalated molecule.

When negatively supercoiled DNA is subjected to centrifugation through the sucrose density gradient with increasing concentrations of ethidium bromide, its sedimentation coefficient first decreases until all negative supercoils are removed, but then increases following the accumulation of positive supercoils. By measuring a critical dye concentration required for the complete relaxation of a negatively supercoiled DNA sample its supercoiling density can be calculated. This is a detailed but laborious approach and it was subsequently replaced by less complex electrophoretic methods. Electrophoretic methods are also based on the difference in shape between supercoiled and relaxed DNA molecules. Circular DNA molecules become more compact with an increase in supercoiling density and migrate faster through an agarose gel than their relaxed counterparts. Consequently, upon separating a mixture of DNA topoisomers in an agarose gel, a ladder of DNA band scan be observed, where the neighbouring bands are chemically identical but differ in number of supercoils by 1. The disadvantage of one-dimensional electrophoresis is that it does not allow the analysis of complex mixtures of DNA topoisomers that might simultaneously include both positively and negatively supercoiled topoisomers of varying densities. This goal can be achieved by using two-dimensional agarose gel electrophoresis. A mixture of DNA topoisomers is first separated in a standard agarose gel. Here the mobility of both

positively and negatively supercoiled DNA topoisomers increases with an increase in their level of supercoiling. Note that upon separation in the first dimension, topoisomers with the same number of supercoils of opposite sign practically co-migrate. To resolve these co-migrating topoisomers, electrophoresis in a second direction, perpendicular to the first one, is performed in the presence of intercalating agent such as ethidium bromide. Since ethidium bromide unwinds DNA, negatively supercoiled topoisomers become less supercoiled and migrate more slowly, while positively supercoiled ones gain extra supercoils and migrate more rapidly.

While the above methods can be used to determine the supercoiling density of circular DNA, they give little information about the shape of supercoiled DNA molecules. This question can be adequately addressed by electron microscopic techniques. Electron microscopy (EM) allows the actual shape of a supercoiled molecule to be seen and the number of crossings per such a molecule to be counted. The disadvantage of conventional electron microscopy is that the DNA conformation may change during sample preparation. Also, it is difficult to determine the sign of a crossing. These problems can be addressed by Cryoelectron microscopy. In this technique, a DNA sample in a water solution is rapidly cooled to $-150\text{ }^{\circ}\text{C}$. As a result, DNA molecules are captured in a thin layer of vitrified water and their three-dimensional images can be obtained (Mirkin 2001).

Recently atomic force microscopy (AFM) imaging approach is used to study the DNA topology. This method allows to lead a better understanding of the behaviour of few kilo base-sized DNA and of the physics governing it at the short-length scales (Allen et al. 1997). AFM was successfully applied to structural studies of numerous biological molecules and their complexes (Arscott and Bloomfield 1992; Engel 1991; Bustamante et al. 1994; Hansma and Hoh 1994; Lyubchenko et al. 1995). A lot of studies demonstrate that AFM may be the method of choice for structural studies of supercoiled DNA. Analyses using AFM have shown that ionic conditions can dramatically change the overall shape of supercoiled DNA so at conditions close to physiological supercoiled DNA adopts a compact plectonemic configuration with close juxtaposition of DNA segments in the loops (Lyubchenko et al. 1997; Shlyakhtenko et al. 2000). Despite its importance, however, much about supercoiled DNA, especially positively supercoiled DNA, remains unknown.

Recently, Irobalieva and colleagues used electron cryo-tomography together with biochemical analyses to investigate structures of individual purified DNA minicircles (336 bp) with defined degrees of supercoiling. Their results reveal that negative and positive minicircles adopt a surprisingly wide distribution of conformations. Moreover, they find striking differences in how the topoisomers handle torsional stress. As negative supercoiling increases, bases are increasingly exposed. Molecular dynamics simulations independently confirm the conformational heterogeneity and provide atomistic insight into the flexibility of supercoiled DNA (Irobalieva et al. 2015).

1.3 Understanding Reverse Gyrase function

Reverse gyrase is a unique DNA topoisomerase that can introduce positive supercoils into DNA (Forterre et al. 1985; Nakasu and Kikuchi 1985; Nadal 2007; Lulchev and Klostermeier 2014; Visone et al. 2014) and it is exclusively found in hyperthermophiles (Brochier-Armanet and Forterre 2007; Forterre 2002), suggesting an important function at high temperatures. Reverse gyrase is a monomeric type IA topoisomerase of 130-kDa and its structure consists of two domains (Confalonieri et al. 1993): a carboxyl terminal domain with topoisomerase activity and an amino terminal domain with an ATP-binding site which presents high similarity to helicase, although neither the whole enzyme nor the isolated helicase-like domain shows helicase activity (Déclais et al. 2000). A crystal structure of the full-length reverse gyrase (Rodríguez and Stock 2002) and more recent structures with additional features (Rudolph et al. 2013) all support the basic two-domain construct. All reverse gyrases isolated and characterized to date catalyze ATP-dependent positive supercoiling of DNA *in vitro*. Although this reaction is considered the hallmark reaction of reverse gyrase, it is unclear whether positive DNA supercoiling is its *in vivo* function (Lulchev and Klostermeier 2014). Positive supercoiling is expected to protect DNA from denaturation at the growth temperatures of hyperthermophiles. A reverse gyrase knockout strain of *Thermococcus kodakaraensis* is viable, but thermosensitive (Atomi et al. 2004), while in the crenarchaeon *Sulfolobus islandicus* both topR1 and topR2 genes were recently demonstrated to be essential (Zhang et al. 2013). However it was demonstrated that positive supercoiling does not efficiently protect DNA from thermodenaturation or thermodegradation (Marguet and Forterre 1994). Moreover a number of observations support a role of reverse gyrase in DNA protection and repair. Reverse gyrase acts as a DNA renaturase that catalyzes annealing of complementary single-stranded DNA circles (Hsieh and Plank 2006). Reverse gyrase is degraded after treatment of *S. solfataricus* with alkylating agent, parallel to the degradation of genomic DNA (Valenti et al. 2006). Furthermore, reverse gyrase interacts with and inhibits a translesion polymerase in *S. solfataricus* (Valenti et al. 2009). Reverse gyrase is recruited to DNA after ultraviolet irradiation, and functionally interacts with single-strand DNA binding protein, linking reverse gyrase to the cell response to DNA damage (Napoli et al 2004 and 2005). The helicase-topoisomerase IA chimeric structure of the reverse gyrase (Nadal 2007; Confalonieri et al. 1993; Jaxel et al. 1996) is reminiscent of the physical and functional interaction between the RecQ-like protein and topoisomerase III (reviewed in Visone et al. 2014). This protein pair is found in Bacteria and Eukarya, and is involved in the DNA repair and recombination needed for genome stability (Gangloff et al. 1994; Harmon et al. 1999; Mankouri and Hickson 2007). It has therefore been suggested that the reverse gyrase in hyperthermophilic archaea has a role in the maintenance of genome stability (Nadal 2007; Perugini et al. 2009).

Recently, Couturier et al. (2014) reported the first quantification of the numbers of molecules of the two reverse gyrases (TopR1 and TopR2) per cell in *S. solfataricus*. Although both enzymes are able to introduce positive supercoils into DNA, TopR2 introduces a higher density of positive supercoils and at a

higher rate than does TopR1. Moreover they exhibit different enzymatic characteristics and in particular different behaviors with respect to temperature: whereas the activity of TopR1 increases progressively with increasing temperature starting from 60°C with maximum positive supercoiling reached at 90°C, TopR2 is not active at high temperature *in vitro* (its optimal temperature is around 70°C), but exhibits a significant positive supercoiling activity at 60°C (Bizard et al. 2011). By contrast, TopR2 was active at 45°C (Couturier et al. 2014).

Studies on reverse gyrase have focused on the catalysis of positive DNA supercoiling by reverse gyrase *in vitro*. The overwinding activity has been demonstrated mainly with gel analyses on a circular DNA substrate where changes in electrophoretic mobility as a result of supercoiling are detected.

Recently Ogawa et al. (2015) observed the activity of purified reverse gyrase from *S. tokodaii* under an optical microscope at 71 °C. In this study a magnetic bead was tethered to a glass surface with linear dsDNA. When reverse gyrase overwound the DNA, the bead rotated, or sank when rotation was constrained. They shown that 30 consecutive mismatched base pairs (a bubble) in DNA serve as a well-defined substrate site for reverse gyrase, warranting the processive overwinding activity down to 50 °C. In a subsequent work Ogawa et al. (2016) inquire how multiple reverse gyrase molecules may collaborate with each other in overwinding one DNA molecule. They demonstrated that rotation with two bubbles was significantly faster compared with one bubble, indicating that overwinding actions are basically additive, but four bubbles did not show further acceleration. When torsional stress in the DNA, determined by the friction, approaches ~ 7 pN_nm (at 71 °C), the overwinding activity of reverse gyrase drops sharply. Multiple molecules of reverse gyrase collaborate additively within this limit.

Although these works highlight the mechanical properties of the reaction of positive supercoiling, many gaps are still present on the role *in vivo* of the reverse gyrase.

1.4 DNA Topoisomerases and their poisoning

DNA topoisomerases are essential enzymes that solve DNA topological problems resulting from strand separation during replication and transcription. Topoisomerases are emerging as important factors in a wide range of metabolic processes in both the nuclear and mitochondrial genomes (Pommier et al. 2016). Topoisomerases are able to covalently modify DNA topology by introducing transient DNA breaks using a transesterification mechanism, which is highly reversible (Champoux 2001; Wang 2002; Pommier et al. 2010; Chen et al. 2013). DNA cleavage by all topoisomerases is accompanied by the formation of a transient phosphodiester bond between a tyrosine residue in the protein and one of the ends of the broken DNA strand; the religation reaction is carried out by attack of the deoxyribose hydroxyl end to the tyrosyl-phosphodiester bond. This cycle of cleavage/religation of DNA strand is realized through the formation of a covalent complex between Top and DNA, usually named “topoisomerase cleavage complex” (Topcc). Topcc is the target of various compounds that act by trapping the complex and,

therefore, they can be employed to treat human cancers. Moreover, failure to complete the catalytic cycle results in trapping of topoisomerase, which remains on DNA and generates DNA breaks (Bunch 2016). The human genome encode for six DNA topoisomerases (TOP1, TOP1mt, TOP2 α , TOP2 β , TOP3 α and TOP3 β), which have both shared and specialized roles. For example, both TOP2 α and TOP2 β relax negatively supercoiled DNA and perform DNA decatenation, but whereas TOP2 α is absolutely required for chromosome segregation, TOP2 β is indispensable for transcription in differentiated, non-dividing cells (Kimura et al. 1994). Furthermore Top2 α is a known biomarker of proliferation, highly expressed in neoplastic cells and undetectable in quiescent cells. Conversely, mature quiescent cells (*e.g.* cardiomyocytes) express Top2 β (present in nucleus and mitochondrion) whereas Top2 α is undetectable (Kondapi et al. 2004).

Drugs poison Top2 cleavage complexes (Top2cc) by two mechanisms: stabilizing the cleavage complex or inhibiting the formation of Top2cc (Nitiss 2009). Anthracyclines are antitumor agents and they act by stabilizing Top2cc and leading to increased levels of Top2–DNA covalent complexes. These agents generate lesions that include DNA strand breaks and protein-DNA covalent complexes.

So far, five members of anthracyclines are approved by USA Food and Drug Administration (FDA) for clinical use: doxorubicin (DOX; Adriamycin®, the most widely used as a first-choice anticancer drug), daunorubicin (Cerubidine®), epirubicin (Ellence®), idarubicin (Idamycin®) and valrubicin (Valstar®) (Minotti et al. 2004; Salvatorelli et al. 2015; Mordente et al. 2009). Structurally, all the anthracyclines are composed of an aglycone (a planar tetracyclic ring system) and a sugar (called “daunosamine”) (Mordente et al. 2012). Anthracyclines antitumor activity is commonly attributed to their ability to intercalate in the DNA double helix, thus poisoning topoisomerases II and ultimately inducing cell death (Nitiss 2009; Pommier et al. 2010).

Despite their large usage in chemotherapy, the molecular mechanisms responsible for anthracycline anticancer activity as well as those underlying cardiotoxicity, the most important side effect of these drugs, remain incompletely understood and still represent a matter of considerable debate and controversy (Minotti et al. 2004; Menna et al. 2011; Mordente et al. 2009; Gewirtz et al. 1999). Although the exact mechanism by which anthracyclines kill tumor cells is still not fully understood, various evidences suggest that Top2 α is the prominent molecular effector of anthracyclines anticancer activity (Minotti et al. 2004; Nitiss 2009; Mordente et al. 2012; Force and Wang 2013). Furthermore it's known that at higher concentrations, anthracyclines act as intercalating agents by interfering with Top2 binding to DNA, thereby suppressing Top2cc formation.

One of the approaches used to overcome the problems related to anthracycline's cardiotoxicity is conjugation with different type of carrier. Among them, oligoarginines are promising candidates.

Oligoarginines belong to the group of cell-penetrating peptides (CPP). They were developed after the discovery that polycationic, Arg-rich peptides origin were able to translocate the cell membrane (Derossi et al. 1994; Fawell et al. 1994). It was found that oligoarginines could be taken up more effectively than

other cationic oligopeptides with equal length; it was also proved that an optimum number of Arg-residue is required for effective internalization (Futaki et al. 2001). Current reviews summarize the present understanding about the mechanism of internalization of CPPs (Patel et al. 2007; Nakase et al. 2008). The mechanism and efficacy of cellular uptake depends also on the physicochemical properties of the cargo attached and on the cell line studied. Delivery of different type of cargos by CPPs was reviewed recently by Hudecz et al. (2005) and Foged et al. (2008). Interestingly, there are few examples for the application of oligoarginines for delivery of small organic compounds.

References

1. Adhya S. Multipartite genetic control elements: communication by DNA loop. *Annu Rev Genet.* 1989; 23:227-50.
2. Allen M.J, Bradbury E.M, Balhorn R. “AFM analysis of DNA–protamine complexes bound to mica. *Nucleic Acids Research*, 1997, 25,112221–2226.
3. Arscott P.G and Bloomfield V.A. Scanning tunneling microscopy of nucleic acids. *Methods Enzymol.* 1992; 211:490-506.
4. Atomi H, Matsumi R, Imanaka T. Reverse gyrase is not a prerequisite for hyperthermophilic life. *J Bacteriol.* 2004 Jul; 186(14): 4829-33.
5. Benjamin H.W and Cozzarelli N.R. DNA-directed synapsis in recombination; slithering and random collision of sites. *Proc. Robert A. Welch Found. Conf. Chem. Res.* 1986; 29: 107-126.
6. Bizard A, Garnier F, Nadal M: TopR2, the Second Reverse Gyrase of *Sulfolobus solfataricus*, Exhibits Unusual Properties. *J Mol Biol* 2011,408:839–849.
7. Brochier-Armanet C and Forterre P. Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea.* 2007 May; 2(2): 83-93.
8. Bunch H. Role of genome guardian proteins in transcriptional elongation. *Febs Lett.*, 2016, 590, 1064-1075.
9. Bustamante C, Erie D.A, Keller D. Biochemical and structural applications of scanning force microscopy. (1994) *Curr. Opin. Struct. Biol.* 3, 750–760.
10. Champoux J. J. DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* 70, 369–413 (2001).
11. Chen S.H, Chan N.L, Hsieh T.S. New mechanistic and functional insights into DNA topoisomerases. *Annu. Rev. Biochem.* 82, 139–170 (2013).
12. Confalonieri F, Elie C, Nadal M, de La Tour C, Forterre P, Duguet M. Reverse gyrase: a helicase-like domain and a type I topoisomerase in the same polypeptide. *Proc Natl Acad Sci U S A.* 1993 May 15; 90(10): 4753-7.
13. Corless S and Gilbert N. Effects of DNA supercoiling on chromatin architecture. *Biophys Rev* (2016) 8:245–258 DOI 10.1007/s12551-016-0210-1.
14. Couturier M, Bizard A.H, Garnier F, Nadal M. Insight into the cellular involvement of the two reverse gyrases from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *BMC Mol Biol.* 2014 Sep 9; 15:18. Doi: 10.1186/1471-2199-15-18.
15. Craigie R and Mizuuchi K. Role of DNA topology in Mu transposition: mechanism of sensing the relative orientation of two DNA segments. *Cell* 1986; 45, 793–800.

16. Déclais AC, Marsault J, Confalonieri F, de La Tour C.B, Duguet M. Reverse gyrase, the two domains intimately cooperate to promote positive supercoiling. *J Biol Chem* 2000, 275(26): 19498–19504.
17. Derossi D, Joliot A.H, Chassaing G, Prochiantz A. The third helix of the Antennapedia homeodomain translocates through biological membranes. *J Biol Chem* 1994, 269, 10444–10450.
18. Drlica K. Control of bacterial DNA supercoiling. *Mol Microbiol.* 1992; 6:425–433.
19. Echols H. Multiple DNA-protein interactions governing high-precision DNA transactions. *Science.* 1986 Sep 5; 233(4768): 1050-6.
20. Engel A. Biological applications of scanning probe microscopes. *Annu Rev Biophys Biophys Chem.* 1991; 20:79-108.
21. Fawell S, Seery J, Daikh Y, Moore C, Chen LL, Pepinsky B, Barsoum J. Tat-mediated delivery of heterologous proteins into cells. *Proc Natl Acad Sci U S A.* 1994 Jan 18; 91(2): 664-8.
22. Foged C, Nielsen HM. Cell-penetrating peptides for drug delivery across membrane barriers. *Expert Opin Drug Deliv.* 2008 Jan; 5(1): 105-17.
23. Force T and Wang Y. Mechanism-based engineering against anthracycline cardiotoxicity. *Circulation*, 2013, 128, 98-100.
24. Forterre P, Mirambeau G, Jaxel C, Nadal M, Duguet M. High positive supercoiling in vitro catalyzed by an ATP and polyethylene glycol-stimulated topoisomerase from *Sulfolobus acidocaldarius*. *Embo J* 1985, 4(8): 2123–2128.
25. Forterre P. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet.* 2002 May; 18(5): 236-7.
26. Futaki S, Suzuki T, Ohashi W, Yagami T, Tanaka S, Ueda K, Sugiura Y. Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *J Biol Chem.* 2001 Feb 23; 276(8): 5836-40.
27. Gangloff S, McDonald J.P, Bendixen C, Arthur L, Rothstein R. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. *Mol Cell Biol* 1994, 14:8391–8398.
28. Gellert M and Nash H. Communication between segments of DNA during site-specific recombination. *Nature.* 1987 Jan 29-Feb 4; 325(6103): 401-4.
29. Gewirtz, D.A. A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics Adriamycin and daunorubicin. *Biochem. Pharmacol.*, 1999, 57, 727-741.
30. Hansma H.G and Hoh J.H. Biomolecular imaging with the atomic force microscope. *Annu Rev Biophys Biomol Struct.* 1994; 23:115-39.

31. Harmon F.G, DiGate R.J, Kowalczykowski S.C. RecQ helicase and topoisomerase III comprise a novel DNA strand passage function: a conserved mechanism for control of DNA recombination. *Mol Cell* 1999, 3:611–620.
32. Hirose S and Matsumoto K. Possible Roles of DNA Supercoiling in Transcription Chapter. In book: *DNA Conformation and Transcription*, pp.138-14, January 2005 DOI: 10.1007/0-387-29148-2_10.
33. Hochschild A. *DNA Topology and Its Biological Effects*. Plainview, NY: Cold Spring Harbor Lab. Press; 1990. pp. 107–116.
34. Hsieh TS, Plank JL. Reverse gyrase functions as a DNA renaturase: annealing of complementary single-stranded circles and positive supercoiling of a bubble substrate. *J Biol Chem*. 2006 Mar 3; 281(9): 5640-7.
35. Hudecz F, Bánóczy Z, Csík G. Medium-sized peptides as built in carriers for biologically active compounds. *Med Res Rev*. 2005 Nov; 25(6): 679-736.
36. Irobalieva R.N, Fogg J.M, Catanese D.J Jr, Sutthibutpong T, Chen M, Barker A.K, Ludtke S.J, Harris S.A, Schmid M.F, Chiu W, Zechiedrich L. Structural diversity of supercoiled DNA. *Nat Commun*. 2015 Oct 12; 6:8440. Doi: 10.1038/ncomms9440.
37. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*. 1961; 3:318–356.
38. Japaridze A, Muskhelishvili G, Benedetti F, Gavriilidou AF, Zenobi R, De Los Rios P, Longo G, Dietler G. Hyperplectonemes: A Higher Order Compact and Dynamic DNA Self-Organization. *Nano Lett*. 2017 Feb 16. Doi: 10.1021/acs.nanolett.6b05294.
39. Jaxel C, Bouthier de la Tour C, Duguet M, Nadal M. Reverse gyrase gene from *Sulfolobus shibatae* B12: gene structure, transcription unit and comparative sequence analysis of the two domains. *Nucleic Acids Res* 1996, 24:4668–4675.
40. Kimura K, Saijo M, Ui M, Enomoto T. Growth state and cell cycle-dependent fluctuation in the expression of two forms of DNA topoisomerase II and possible specific modification of the higher molecular weight form in the M phase. *J. Biol. Chem.*, 1994, 269, 1173-1176.
41. Kondapi A.K, Mulpuri N, Mandraju R.K, Sasikaran B, Subba Rao K. Analysis of age dependent changes of Topoisomerase II alpha and beta in rat brain. *Int. J. Dev. Neurosci.*, 2004, 22, 19-30.
42. Lulchev P and Klostermeier D. Reverse gyrase—Recent advances and current mechanistic understanding of positive DNA supercoiling. *Nucleic Acids Res* 2014, 42(13): 8200–8213.
43. Lyubchenko Y.L, Jacobs B.L, Lindsay S.M, Stasiak A. Atomic force microscopy of nucleoprotein complexes. *Scanning Microsc*. 1995 Sep;9(3):705-24; discussion 724-7.
44. Lyubchenko Y.L, Shlyakhtenko L.S, Aki T, Adhya S. Atomic force microscopic demonstration of DNA looping by GalR and HU. *Nucleic Acids Res*. 1997 Feb 15; 25(4): 873-6.

45. Mankouri H.W, Hickson I.D. The RecQ helicase-topoisomerase III-Rmi1 complex: a DNA structure-specific ‘dissolvasome’? *Trends Biochem Sci* 2007, 32:538–546.
46. Margolin P, Zumstein L, Sternglanz R. et al. The *Escherichia coli* supX locus is topA, the structural gene for DNA topoisomerase I. *Proc Natl Acad Sci USA*. 1985; 82:5437–5441.
47. Marguet E, Forterre P. DNA stability at temperatures typical for hyperthermophiles. *Nucleic Acids Res*. 1994 May 11; 22(9): 1681-6.
48. Menna P, Gonzalez Paz O, Chello M, Covino E, Salvatorelli E, Minotti G. Anthracycline cardiotoxicity. *Expert Opin. Drug Saf.*, 2011.
49. Menzel R, Gellert M. Modulation of transcription by DNA supercoiling: A deletion analysis of the *Escherichia coli* gyrA and gyrB promoters. *Proc Natl Acad Sci USA*. 1987; 84:4185–4189.
50. Meyer S, Beslon G. Torsion-mediated interaction between adjacent genes. *PLoS Comput Biol*. 2014 Sep 4; 10(9): e1003785. Doi: 10.1371/journal.pcbi.1003785. ECollection 2014.
51. Minotti G, Menna P, Salvatorelli E, Cairo G, Gianni L. Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. *Pharmacol. Rev.*, 2004, 56, 185-229.
52. Mirkin S.M DNA Topology: Fundamentals. *Encyclopedia Of Life Sciences / 2001 Nature Publishing Group*.
53. Mordente A, Meucci E, Silvestrini A, Martorana G.E, Giardina B. Anthracyclines and mitochondria. *Adv. Exp. Med. Biol.*, 2012, 942, 385-419.
54. Mordente A, Meucci E, Silvestrini A, Martorana G.E, Giardina B. New developments in anthracycline-induced cardiotoxicity. *Curr. Med. Chem.*, 2009, 16, 1656-1672.
55. Nadal M. Reverse gyrase: An insight into the role of DNA-topoisomerases. *Biochimie* 2007, 89(4): 447–455.
56. Nakase I, Takeuchi T, Tanaka G, Futaki S. Methodological and cellular aspects that govern the internalization mechanisms of arginine-rich cell-penetrating peptides. *Adv Drug Deliv Rev*. 2008 Mar 1; 60(4-5): 598-607.
57. Nakasu S and Kikuchi A. Reverse gyrase, ATP-dependent type I topoisomerase from *Sulfolobus*. *Embo J*1985, 4(10): 2705–2710.
58. Napoli A, Valenti A, Salerno V, Nadal M, Garnier F, Rossi M, Ciaramella M. Reverse gyrase recruitment to DNA after UV light irradiation in *Sulfolobus solfataricus*. *J. Biol. Chem*. 2004, 279, 33192–33198.
59. Nitiss J.L. Targeting DNA topoisomerase II in cancer chemotherapy. *Nat. Rev. Cancer*, 2009, 9, 338-350.
60. Ogawa T, Sutoh K, Kikuchi A, Kinoshita K Jr. Torsional stress in DNA limits collaboration among reverse gyrase molecules. *Febs J*. 2016 Apr; 283(8): 1372-84. doi: 10.1111/febs.13675.

61. Ogawa T, Yogo K, Furuike S, Sutoh K, Kikuchi A, Kinoshita K Jr. Direct observation of DNA overwinding by reverse gyrase. *Proc Natl Acad Sci U S A*. 2015 Jun 16; 112(24): 7495-500. Doi: 10.1073/pnas.1422203112.
62. Patel L N, Zaro J L, and Shen W C. Cell penetrating peptides: intracellular pathways and pharmaceutical perspectives. *Pharm Res*. 2007 Nov; 24(11): 1977-92.
63. Perugini G, Valenti A, D'amaro A, Rossi M, Ciaramella M. Reverse gyrase and genome stability in hyperthermophilic organisms. *Biochem Soc Trans* 2009, 37:69–73.
64. Pommier Y, Leo E, Zhang H, Marchand C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem. Biol.*, 2010, 17, 421-433.
65. Pommier Y, Sun Y, Huang S.N and Nitiss J L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev.* Vol 17; November 2016 doi: 10.1038/nrm.2016.111
66. Pruss GJ, Drlica K. DNA supercoiling and prokaryotic transcription. *Cell*. 1989; 56:521–523.
67. Rodríguez A.C and Stock D. Crystal structure of reverse gyrase: Insights into the positive supercoiling of DNA. *Embo J* 2002, 21(3):418–426.
68. Rudd KE, Menzies R. His operons of *Escherichia coli* and *Salmonella typhimurium* are regulated by DNA supercoiling. *Proc Natl Acad Sci USA*. 1987; 84:517–521.
69. Rudolph M.G, del Toro Duany Y, Jungblut S.P, Ganguly A, Klostermeier D. Crystal structures of *Thermotoga maritima* reverse gyrase: Inferences for the mechanism of positive DNA supercoiling. *Nucleic Acids Res* 2013, 41(2): 1058–1070.
70. Salvatorelli E, Menna P, Cantalupo E, Chello M, Covino E, Wolf F.I, Minotti G. The concomitant management of cancer therapy and cardiac therapy. *Biochim. Biophys. Acta*, 2015, 1848, 2727-2737.
71. Schleif R. DNA looping. *Annu Rev Biochem*. 1992; 61:199-223.
72. Serre M.C and Duguet M. Enzymes that cleave and religate DNA at high temperature: the same story with different actors. *Prog Nucleic Acid Res Mol Biol* 2003, 74:37–81.
73. Shlyakhtenko L.S, Hsieh P, Grigoriev M, Potaman V.N, Sinden R.R, Lyubchenko Y.L. A cruciform structural transition provides a molecular switch for chromosome structure and dynamics. *J Mol Biol*. 2000 Mar 10; 296(5): 1169-73.
74. Sinden R.R. *DNA Structure and Function*. Academic Press, San Diego, 1994, pp. 95–128.
75. Sobetzko P. Transcription-coupled DNA supercoiling dictates the chromosomal arrangement of bacterial genes. 1514–1524 *Nucleic Acids Research*, 2016, Vol. 44, No. 4 Published online 17 January 2016 doi: 10.1093/nar/gkw007.
76. Summers D.W, Ernst C, Spengler S.J, Cozzarelli N.R. Analysis of the mechanism of DNA recombination using tangles. *Q Rev Biophys*. 1995 Aug; 28(3): 253-313.

77. Valenti A, Napoli A, Ferrara M.C, Nadal M, Rossi M, Ciaramella M. Selective degradation of reverse gyrase and DNA fragmentation induced by alkylating agent in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.* 2006, 34, 2098–2108.
78. Valenti A, Perugino G, Nohmi T, Rossi M, Ciaramella M. Inhibition of translesion DNA polymerase by archaeal reverse gyrase. *Nucleic Acids Res.* 2009, 37, 4287–4295.
79. Visone V, Vettone A, Serpe M, Valenti A, Perugino G, Rossi M, Ciaramella M. Chromatin structure and dynamics in hot environments: architectural proteins and DNA topoisomerases of thermophilic archaea. *Int J Mol Sci.* 2014 Sep 25;15(9):17162-87. doi: 10.3390/ijms150917162.
80. Wang J.C. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* 3, 430–440 (2002).
81. Wang J.C. DNA topoisomerases. *Annu Rev Biochem.* 1985; 54: 665-97.
82. Zechiedrich E.L and Osheroff N. Eukaryotic topoisomerases recognize nucleic acid topology by preferentially interacting with DNA crossovers. *Embo J.* 1990 Dec; 9(13):4555-62.
83. Zhang C, Tian B, Li S, Ao X, Dalgaard K, Gökce S, Liang Y, She Q: Genetic manipulation in *Sulfolobus islandicus* and functional analysis of DNA repair genes. *Biochem Soc Trans* 2013, 41:405–410.

2. Papers

2.1 Massively Systematic Transcript End Readout (MASTER): Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields

Irina O. Vvedenskaya^{1,2,*}, Yuanchao Zhang^{1,3,*}, Seth R. Goldman¹, Anna Valenti⁴, Valeria Visone⁴, Deanne M. Taylor^{1,3}, Richard H. Ebright^{2,5}, and Bryce E. Nickels^{1,2}

¹Department of Genetics, Rutgers University, Piscataway, New Jersey 08854

²Waksman Institute, Rutgers University, Piscataway, New Jersey 08854

³Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA 19041

⁴Institute of Biosciences and Bioresources, National Research Council of Italy, Via P. Castellino 111, Naples 80131, Italy

⁵Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854

Published in final edited form as:

Mol Cell. 2015 December 17; 60(6): 953–965. Doi: 10.1016/j.molcel.2015.10.029.

SUMMARY

We report the development of a next-generation sequencing-based technology that entails construction of a DNA library comprising up to at least 4^7 (~16,000) bar-coded sequences, production of RNA transcripts, and analysis of transcript ends and transcript yields ("massively systematic transcript end readout," MASTER). Using MASTER, we define full inventories of transcription start sites ("TSSomes") of *Escherichia coli* RNA polymerase for initiation at a consensus core promoter *in vitro* and *in vivo*, we define the TSS-region DNA-sequence determinants for TSS selection, reiterative initiation ("slippage synthesis"), and transcript yield, and we define effects of DNA topology and NTP concentration. The results reveal that slippage synthesis occurs from the majority of TSS-region DNA sequences and that TSS-region DNA sequences have profound, up to 100-fold, effects on transcript yield. The results further reveal that TSSomes depend on DNA topology, consistent with the proposal that TSS selection involves transcription-bubble expansion ("scrunching") and transcription-bubble contraction ("anti-scrunching").

INTRODUCTION

Defining the mechanism of any process involving protein-nucleic acid interactions requires a full understanding of the relationship between nucleic acid sequence and functional output. Often such studies require the construction and analysis of many individual variants of a given DNA sequence. However, for processes involving extensive protein-nucleic acid interactions, a comprehensive analysis of the functional output derived from all sequence variants is not feasible with conventional approaches that test individual sequence variants on a one-by-one basis. Instead, such analyses require approaches

that facilitate the parallel interrogation of thousands of individual sequence variants en masse (Melnikov et al., 2012; Patwardhan et al., 2012; Heyduk and Heyduk, 2014; Heyduk and Heyduk, 2015).

The process of transcription involves extensive interactions between RNA polymerase (RNAP) and a double-stranded DNA template. Each stage of transcription, and each reaction carried out by RNAP during transcription, is affected by the sequence context of the DNA template (Larson et al., 2011; Wang and Greene, 2011; Dangkulwanich et al., 2014; Ruff et al., 2015; Washburn and Gottesman, 2015). While structural studies have revealed some of the RNAP-nucleic acid interactions that modulate transcription (Murakami, 2015), a quantitative understanding of how DNA sequence influences transcription output requires comprehensive knowledge of RNAP activity in all sequence contexts. Toward achievement of this goal, we report the development of a technology that facilitates the comprehensive analysis of the relationship between nucleic acid sequence and functional output during transcription for *Escherichia coli* RNAP on a scale previously inaccessible. In particular, we report the development of an experimental platform that enables the parallel functional analysis of at least 4^7 (~16,000), and potentially at least 4^{10} (~1,000,000), distinct DNA template sequence variants *in vitro* and *in vivo* ("massively systematic transcript end readout," MASTER). We apply MASTER to the analysis of the determinants and mechanism of transcription initiation. Specifically, by use of MASTER we (i) perform a comprehensive analysis of the DNA-sequence determinants for transcription start site (TSS) selection, (ii) perform a comprehensive analysis of DNA-sequence determinants for reiterative initiation (a non-standard form of transcription initiation also termed "slippage synthesis"), (iii) perform a comprehensive analysis of the relationship between TSS-region sequence and transcript yield, and (iv) define the effects of DNA topology and alterations in nucleoside triphosphate (NTP) concentrations on TSS-selection, slippage synthesis, and transcript yield. Our results provide a full description of the relationship between TSS- region sequence and transcriptional output from a consensus core promoter for *E. coli* RNAP both *in vitro* and *in vivo*. Our results further demonstrate the broad utility of MASTER for analysis of biological processes that involve extensive protein-nucleic acid interactions.

DESIGN

To overcome limitations of conventional approaches for the analysis of effects of sequence variation on transcription, we sought to develop a technology that would enable a comprehensive analysis of the relationship between sequence and transcription output. Thus, we developed a next-generation sequencing-based approach (Figure 1) that entails construction of a template library that contains up to 4^7 (~16,000), and potentially up to 4^{10} (~1,000,000), barcoded sequences (Figure S1A,B), production of RNA transcripts from the template library *in vitro* or *in vivo*, and analysis of transcript ends and transcript yields. To generate a template library, we synthesize an oligodeoxyribonucleotide containing a first randomized sequence spanning the region of interest (Figure 1, top, green) and a second, longer randomized sequence that serves as a barcode (Figure 1, top, blue). The oligodeoxyribonucleotide is converted to double-stranded DNA by use of PCR, the double-stranded DNA is ligated with a vector

backbone, the resulting recombinant plasmid DNA is used to transform cells, and recombinant plasmid DNA is isolated from 10^6 - 10^7 transformants (Figure 1, middle). The resulting library is amplified by emulsion PCR and sequenced to define each unique sequence and thereby to associate each first randomized sequence with its corresponding second randomized sequence that serves as its barcode. This procedure provides a “self-assembling barcode,” in which for each DNA molecule in the library, the first randomized sequence in the library is associated with a known corresponding second randomized sequence. RNA produced from the library, either in an *in vitro* transcription reaction or *in vivo*, is isolated and subjected to deep sequencing in a manner that enables the identification of the sequence of transcript ends and the sequence of the barcode within each transcript (Figure 1, bottom). We term this approach “massively systematic transcript end readout, MASTER.”

RESULTS

Application of MASTER for analysis of transcription initiation

During transcription initiation, the RNAP holoenzyme binds to promoter DNA, unwinds a turn of promoter DNA to form an RNAP-promoter open complex containing an unwound “transcription bubble,” and selects a transcription start site (TSS) (Ruff et al., 2015). Some progress has been made to define the DNA sequence determinants for TSS selection (Aoyama and Takanami, 1985; Sorensen et al., 1993; Jeong and Kang, 1994; Liu and Turnbough, 1994; Walker and Osuna, 2002; Lewis and Adhya, 2004). However, a complete description of how the sequence of the TSS-region of a given promoter determines TSS selection, and the impact of the sequence of the TSS-region on yields of full-length transcripts, has not been provided. We therefore sought to employ MASTER to perform a comprehensive analysis of the effects of TSS-region sequences on functional output during transcription initiation for *E. coli* RNAP.

We constructed a MASTER library containing 4^7 (~16,000) sequences at the TSS-region, i.e. positions 4–10 base pairs (bps) downstream of the -10 element, of a consensus *E. coli* σ^{70} -dependent promoter (Figures 1, S1C). We generated RNA from the library in transcription reactions performed *in vitro* using non-supercoiled DNA templates, transcription reactions performed *in vitro* using negatively supercoiled DNA templates, and transcription of negatively supercoiled DNA templates *in vivo*. The RNA transcripts were analyzed by high-throughput sequencing of RNA 5' ends (5' RNA-seq) to identify the sequence of the RNA 5' end and to identify the sequence of the barcode, which defined the identity of the template that produced the RNA. The sequences of RNA 5' ends generated from a given template in the library were used to define the TSS position(s) from the template and the total number of sequencing reads generated from the template were used to define transcript yields from the template (Figure 1, bottom).

MASTER analysis of sequence determinants of TSS selection

To analyze TSS selection we considered only sequencing reads with 5' end sequences that precisely matched the sequence of the TSS region from which the transcript was generated “matched RNAs”. For each TSS-region sequence we calculated the number of matched RNAs emanating from each position 4–

10 bps downstream of the -10 element (hereafter termed positions 4, 5, 6, 7, 8, 9 and 10). Next, we calculated the percentage of reads derived from matched RNAs emanating from each position (X) within the TSS region [$\%TSS_X = 100 * (\# \text{ reads at position X}) / (\# \text{ reads at all positions 4-10})$], and used the $\%TSS_X$ values to calculate a mean TSS for each sequence variant. The results define the full TSS inventories ("TSSomes") from a consensus core promoter for non-supercoiled templates *in vitro*, negatively supercoiled templates *in vitro*, and negatively supercoiled templates *in vivo* (Tables S1–S7).

MASTER analysis of TSS selection using non-supercoiled linear DNA

Averaging the $\%TSS$ at positions 4–10 observed for all TSS-sequence variants in reactions performed using non-supercoiled linear DNA templates (Table S1) shows, consistent with previous analyses of individual promoters (Aoyama and Takanami, 1985; Sorensen et al., 1993; Jeong and Kang, 1994; Liu and Turnbough, 1994; Walker and Osuna, 2002; Lewis and Adhya, 2004), that TSS selection is determined by position (Figure 2A) (range = positions 6–10; mean = 7.36 bp downstream of -10 element; mode = 7 bp downstream of -10 element). The results further show the order of preference for TSS selection for non- supercoiled linear templates *in vitro* is position $7 > 8 > 6 \sim 9 > 10$ (Figures 2A,B, S2A; $p < 0.001$).

To investigate sequence determinants for TSS selection we sorted TSS-region sequences on the basis of the identities of the bases at positions 6–10. As prior analyses of TSS selection have noted a bias for purine at the TSS position (Maitra and Hurwitz, 1965; Jorgensen et al., 1969; Hawley and McClure, 1983), we first determined the $\%TSS$ for promoter variants that specify use of a purine (R) TSS or pyrimidine (Y) TSS at positions 6–10. The results show a strong preference for R over Y at each TSS position (Figures 2B, S2B; $p < 0.001$). Further analysis of TSS region variants carrying an A, G, C, or T at each TSS position revealed the order of preference as $G > A > C \sim T$ (Figures 2B, S2C; $p < 0.001$). Prior analyses of TSS selection also noted a preference for initiation at YTSS-1RTSS sequence motifs (Shultzaberger et al., 2007). To investigate the influence of YTSS-1RTSS sequences on TSS selection we determined the $\%TSS$ for promoter variants carrying YTSS-1RTSS or RTSS-1RTSS. The results show that YTSS-1RTSS is preferred over RTSS-1RTSS at each TSS position (Figures 2B, S2D; $p < 0.001$).

We next identified TSS-region sequences that yielded the highest $\%TSS$ at each of the five positions within the TSS range (Figure 2C). For each TSS position, sequences that favor the highest $\%TSS$ exhibit a strong preference for YTSS-1RTSS. In addition, for TSS position 9 there is a strong preference for TTSS-1, while for TSS position 10, the sequences with the highest $\%TSS$ show a preference for a TTSS-2TTSS-1.

MASTER analysis of TSS selection using negatively supercoiled DNA

For analysis of the effects of DNA topology on TSS selection we compared results obtained *in vitro* using either a non-supercoiled linear DNA template (Table S1) or a relaxed circular DNA template (Table S6) with results obtained using a negatively supercoiled DNA template *in vitro* (Tables S2, S7)

and *in vivo* (Table S3). Analysis of the %TSS at positions 4–10 observed for all TSS-sequence variants in reactions performed using negatively supercoiled DNA templates (Figure 3A,C) shows that the range of TSS positions (positions 6–10) and modal TSS (position 7) was identical to the range of TSS positions and modal TSS observed using a non-supercoiled DNA template (Figure 2A).

In addition, the sequence determinants for TSS selection observed using negatively supercoiled DNA templates (Figure S3) were identical to those observed using non-supercoiled DNA templates (Figure 2B,C). In particular, R is favored over Y at each TSS position, the order of TSS preference for each base is $G > A > C \sim T$, and YTSS-1RTSS is preferred over RTSS-1RTSS at each TSS position (Figure S2A–D; $p < 0.001$). Furthermore, the TSS-region sequences that yielded the highest %TSS at each of the five TSS positions within the TSS range were similar on negatively supercoiled DNA templates and non-supercoiled DNA templates (Figures 2C, S3).

However, there were also notable differences between results obtained using negatively supercoiled DNA templates and non-supercoiled DNA templates. First, while the %TSS observed at position 6 and position 9 were similar with non-supercoiled DNA (Figures 2A, 2B, S4A), the %TSS at position 9 was ~3-times larger than the %TSS at position 6 with negatively supercoiled DNA (Figures 3A, 3C, S3, S4B). Second, comparison of the mean TSS observed with negatively supercoiled DNA (Figure 3A,C) with the mean TSS observed using non-supercoiled DNA (Figure 2A) reveals TSS distributions with negatively supercoiled DNA (mean = 7.65 bp downstream of –10 element with negatively supercoiled DNA *in vitro*; mean = 7.59 bp downstream of –10 element with negatively supercoiled DNA *in vivo*) are shifted downstream relative to TSS distribution with non-supercoiled DNA (mean = 7.36 bp downstream of –10 element). In addition, ~95% of TSS-region sequence variants exhibited an increase in the mean TSS with negatively supercoiled DNA templates compared with a non-supercoiled DNA template (Figures 3B, 3D, S4C).

Analysis of the effects of TSS-region sequence on sensitivity to topology revealed that TSS-region sequences carrying R at positions 7 and 8 are less susceptible to topology effects, while TSS-region sequences carrying a Y at positions 7 and 8 are more susceptible to topology effects (Figures 3E, S2E, S4D; $p < 0.001$). In addition, TSS-region sequences that contained YR at positions 6/7 are less susceptible to topology effects, while TSS-region sequences carrying a RY at positions 6/7 are more susceptible to topology effects (Figures 3E, S2E, S4D; $p < 0.001$). Furthermore, identification of the TSS-region sequences that exhibited the highest difference in mean TSS in a comparison of negatively supercoiled DNA and non-supercoiled DNA revealed that TSS-region sequences that contained RYYR at positions 6–9 or RYYYYR at positions 6–10 were highly susceptible to topology dependent changes in TSS selection (Figures 3E, 3F, S2E, S4D, S4E; $p < 0.001$).

The results obtained from our MASTER analysis of TSS selection on non-supercoiled and negatively supercoiled DNA indicate that DNA topology is a determinant of TSS selection and reveal TSS-region sequence determinants that confer sensitivity or resistance to topology-dependent changes in TSS

selection for a consensus core promoter. In addition, a comparison of the results obtained using negatively supercoiled DNA *in vitro* (Figures 3A, S3A, S4B) with results obtained using negatively supercoiled DNA *in vivo* (Figures 3C, S3B) show that TSS selection with negatively supercoiled DNA *in vitro* accurately recapitulates TSS selection with negatively supercoiled DNA *in vivo*, suggesting that all determinants for TSS selection for a consensus core promoter are contained within an *in vitro* transcription reaction (i.e., RNAP, DNA, and NTPs) and that no other determinants or factors have major effects on global TSS distributions *in vivo* (Figure 3C) or on global sequence determinants for TSS selection *in vivo* (Figure S3B).

MASTER analysis of effects of NTP concentrations on TSS selection

To analyze the effects of NTP concentrations on TSS selection we compared results obtained from the analysis of *in vitro* reactions performed in the presence of saturating NTPs (2.5 mM NTPs:Mg²⁺) (Table S4) with results obtained from the analysis of *in vitro* reactions performed in the presence of non-saturating NTPs (0.1 mM) (Table S5). Results show that the range of TSS positions (6–10) and modal TSS (position 7) are identical at saturating and non-saturating NTP concentrations (Figure 4A,B) and identical to the range and mode observed in the analysis of non-supercoiled and supercoiled DNA templates *in vitro* at 1 mM NTPs and the analysis of supercoiled DNA templates *in vivo* (Figures 2A, 3A, 3C, S4A, S4B). Comparison of the sequence determinants for TSS selection at saturating and non-saturating NTP concentrations revealed that preference for an R TSS over a Y TSS increases at non-saturating NTP concentrations, preference for a G TSS increases at non-saturating NTP concentrations, and preference for YTSS-1RTSS over RTSS-1RTSS increases at non-saturating NTP concentrations (Figures 4C, 4D, S2A–D).

Comparison of the mean TSS observed at saturating and non-saturating NTP concentrations (Figure 4A,B) reveals the overall TSS distribution at non-saturating NTP concentrations (mean = 7.50 bp downstream of –10 element) is shifted slightly downstream relative to the overall TSS distribution at saturating NTP concentrations (mean = 7.38 bp downstream of –10 element). Comparison of individual TSS-sequence variants revealed that a majority of TSS-sequence variants exhibited a change in the mean TSS at non-saturating NTP concentrations compared with saturating NTP concentrations (Figure 4E). TSS-region sequences carrying an R at position 7 and position 8 are less susceptible to alterations in TSS selection in response to alterations in NTP concentrations, while TSS-region sequences carrying a Y at position 7 and position 8 are more susceptible to alterations in TSS selection in response to alterations in NTP concentrations (Figures 4F, S2E; $p < 0.001$). In addition, TSS-region sequences that contained YR at positions 6/7 are less susceptible to effects of alterations in NTP concentrations, while TSS-region sequences carrying a RY at positions 6/7 are more susceptible to effects of alterations in NTP concentrations (Figures 4F, S2E; $p < 0.001$). Furthermore, TSS-region sequences that contained RYYR at positions 6–9 or RYYYR at positions 6–10 were highly susceptible to NTP-concentration dependent changes in TSS selection (Figures 4F, S2E; $p < 0.001$). Strikingly, the patterns observed in the

comparison of sequences that are more or less susceptible to NTP-concentration dependent changes in TSS selection (Figure 4F) were similar to the patterns observed in the comparison of sequences that are more or less susceptible to topology dependent changes in TSS selection (Figures 3E, S4D).

MASTER analysis defines the extent of productive slippage synthesis during transcription initiation

During the standard pathway of transcription initiation, in each nucleotide addition step, RNAP translocates relative to the DNA and RNA, and the DNA template strand and the 3' end of the RNA product remain in register (Figure 5A). However, during initial transcription RNAP can enter into an alternative pathway of transcription termed "slippage synthesis" (Jacques and Kolakofsky, 1991; Turnbough and Switzer, 2008; Turnbough, 2011). In slippage synthesis, RNAP does not translocate relative to the DNA and RNA, and, instead, the RNA product slips upstream relative to the DNA template strand, establishing a new register of DNA and RNA (Figure 5B). Slippage synthesis can occur in multiple cycles, including very large numbers of cycles. Accordingly, slippage synthesis is also referred to as "reiterative transcription initiation."

RNAs produced as a consequence of slippage synthesis either can be released from the initial transcribing complex ("non-productive slippage synthesis") or can be extended to yield full-length RNA products ("productive slippage synthesis"). Full-length RNA products generated by productive slippage synthesis typically contain at least one 5' nucleotide that does not match the sequence of the DNA template ("RNA/DNA difference"; Figure 5B).

Therefore, to identify products of productive slippage synthesis in our analysis of transcription output, we enumerated reads that carried at least one RNA/DNA difference (Tables S1–S7) and, since slippage synthesis can occur in multiple cycles, we also enumerated reads with 5' ends that were up to five bases longer than position 4, i.e. the first position of the randomized TSS-region ("L reads" in Tables S1–S7).

To assess the extent of productive slippage synthesis, we analyzed the total transcription output from each TSS-region sequence variant for transcripts predicted to be generated by productive slippage synthesis.

First, we assessed the extent of productive slippage synthesis at promoters that contain homopolymeric repeat sequences that start at the TSS (i.e., T_n , A_n , G_n , and C_n , where $n > 1$ and where the sequence starts at the TSS). Such sequences potentially are expected to facilitate productive slippage synthesis, in increments of one base pair, since homopolymeric repeat sequences allow slippage to occur, in increments of one base pair, with the net loss of only one RNA-DNA base pair (Figure 5B). Products of standard synthesis and productive slippage synthesis from homopolymeric repeat sequences were distinguished by the absence or presence, respectively, of at least one RNA/DNA difference. For example, for the TSS- region sequence AAT, which carries an A_2 homopolymeric repeat starting at the underlined TSS base, RNAs generated by standard synthesis would have the fully templated 5'-end sequence AAU-, whereas potential RNAs generated by productive slippage synthesis would have 5'-end

sequences AAAU-, AA AAU-, AAAAAU-, etc. that carry one or more 5'- RNA/DNA difference. We calculated the percentage of RNAs from the homopolymeric tract that are produced by slippage [% slippage = 100(#slippage reads) / (#slippage reads + #standard reads)]. The results indicate that: (i) slippage occurs at promoters that contain T_n, A_n, C_n, and G_n homopolymeric repeats at TSS positions 6, 7, 8, and 9 (Figure 5C); (ii) % slippage is especially high for T_n and A_n homopolymeric repeats (up to >80%; Figure 5C); (iii) % slippage increases in all cases as the length of a T_n, A_n and G_n homopolymeric repeat increases and increases in many cases as the length of a C_n homopolymeric repeat increases (Figure 5C); (iv) the number of nucleotides added to the RNA 5' end by repeated cycles of slippage increases as the length of a T_n, A_n and G_n homopolymeric repeat increases (Tables S1–S7); and (v) the number of nucleotides added to the 5' end by repeated cycles of slippage can be strikingly long (up to at least 8 for T_n and A_n homopolymeric repeats; Tables S1–S7).

Next, we assessed the extent of productive slippage synthesis at promoters that contain homopolymeric repeat sequences that start one base downstream of the TSS. Such sequences also potentially are expected to facilitate productive slippage synthesis, in increments of one nucleotide, since such sequences also allow slippage to occur, in increments of one nucleotide, with the net loss of only one RNA-DNA base pair. Products of standard synthesis and potential products of productive slippage synthesis again were distinguished by the absence or presence, respectively, of at least one RNA/DNA difference. The results indicate that (i) slippage occurs at promoters that contain T_n, A_n, C_n, and G_n homopolymeric repeat sequences that start one base downstream of a TSS at positions 6, 7, and 8 (Figure S5A); (ii) % slippage is especially high for T_n and A_n homopolymeric repeats (Figure S5A); (iii) % slippage increases in all cases as the length of a T_n, A_n and G_n homopolymeric repeat increases and increases in many cases as the length of a C_n homopolymeric repeat increases (Figure S5A); (iv) the number of nucleotides added to the RNA 5' end by repeated cycles of slippage increases as the length of a T_n, A_n and G_n homopolymeric repeat increases (Tables S1–S7); and (v) the number of nucleotides added to the 5' end by repeated cycles of slippage can be strikingly long (up to at least 8 for T_n and A_n homopolymeric repeats; Tables S1–S7).

Next, we assessed the extent of productive slippage synthesis at promoters that do not contain homopolymeric repeat sequences starting at the TSS. Such sequences potentially may allow productive slippage synthesis in increments of two nucleotides. Products of standard synthesis and potential products of productive slippage synthesis were distinguished, in this case, by extension of the RNA 5'-end by increments of two nucleotides. For example, for the TSS-region sequence AGAT, RNAs generated by standard synthesis would have the fully templated 5'-end sequence AGAU-, whereas potential RNAs generated by productive slippage synthesis in increments of two nucleotides would have the 5'-end sequences AGAGAU-, AGAGAGAU-, AGAGAGAGAU-, etc. The results indicate that slippage in increments of two nucleotides occurs for promoters that lack homopolymeric repeat

sequences at the TSS (~0.1% to ~1% slippage in increments of two base pairs; Figure S5C). The results further indicate that slippage in increments of two base pairs is especially high for the TSS-region dinucleotide sequences TG, TA, CG, and AG (Figure S5C), and that the local sequence context of the TSS-region dinucleotide sequence affects % slippage. Thus, for most cases, when the nucleotide preceding the TSS-region dinucleotide sequence matches the second nucleotide of the TSS-region dinucleotide sequence, or when the nucleotide following the TSS-region dinucleotide sequence matches the first nucleotide of the TSS-region dinucleotide sequence, the % slippage increases (Figure S5C). More broadly, when sequences preceding or following the TSS-region dinucleotide sequence (ab) create a dinucleotide repeat pattern of the form b-ab, ab-ab, ab-a, b-ab-a, or ab-ab-a, % slippage is especially higher (Figure S5C). We infer that slippage in increments of two nucleotides occurs, is widespread, and is facilitated by upstream or downstream extensions of the TSS region dinucleotide sequence that increase complementarity between the RNA product and the DNA template strand upon slippage in increments of two nucleotides.

Prompted by the above results, we re-analyzed published RNA-seq data obtained *in vivo* in *E. coli* (Thomason et al, 2015), assessing levels of productive slippage synthesis for chromosomal promoters *in vivo* in *E. coli* (Figure S5B, C). We find extents of slippage for chromosomal promoters containing homopolymeric repeats at the TSS (Figure S5B, top) or one base pair downstream of the TSS (Figure S5B, bottom) similar to those observed for a consensus core promoter by MASTER (Figures 5C, S5A). In addition, we find slippage in increments of two nucleotides at chromosomal promoters containing non-repeat TSS-region dinucleotide sequences (Figure S5C).

Finally, we assessed the contribution of productive slippage synthesis to the total transcription output for each TSS-region sequence variant. We divided the total number of sequencing reads derived from productive slippage synthesis by the total number of sequencing reads, combining data for all analyzed experimental conditions. The results indicate, surprisingly, that productive slippage synthesis is detectable ($\geq 0.5\%$) for the majority of TSS-region sequences (~52%; 8,590 of 16,384), occurs at moderate levels ($\geq 5\%$) for ~18% of TSS-region sequences (3,007 of 16,384), and accounts for the majority of the total transcription output ($\geq 50\%$), for a large number of TSS-region sequences (3.1%; 509 out of 16,384).

MASTER analysis reveals effects of TSS-region sequences on transcript yield

To document effects of TSS-region sequence on relative transcript yield, we divided the read count representing the total transcription output for each TSS-region sequence variant by the relative number of DNA templates that carried the sequence variant. We refer to the value obtained as the “relative expression” of each TSS-region sequence variant (Tables S1– S7). A comparison of the relative expression values allowed us to determine the influence of sequence variation in the TSS-region on the range of expression observed for a given experimental condition (Figures 6A, 6B, S6).

For experiments performed *in vitro* using non-supercoiled linear DNA or negatively supercoiled DNA (at 1 mM NTPs) TSS-region sequence variation led to a ~40-fold range of relative expression (Figure S6A, B). For experiments performed with negatively supercoiled DNA *in vivo* we found that TSS-region sequence variation led to a > 100-fold range of relative expression (Figure S6C). However, interpreting the effects of TSS-region sequences on transcript yields *in vivo* is complicated by the potential contribution of sequence variation at the RNA 5' end to the stability of full-length transcripts. Thus, we infer that the analysis of effects of TSS-region sequence variation on relative transcript yield *in vitro* provides a more accurate measure of the true impact of TSS-region sequence variation on transcription output. We therefore conclude that DNA topology does not exert a significant global impact on the relationship between TSS-region sequence and the range of expression.

We next compared the effect of NTP concentrations on the range of expression. For assays performed *in vitro* at saturating NTP concentrations, TSS-region sequence variation led to a ~14-fold range of relative expression (Figure 6A). In contrast, for assays performed *in vitro* at non-saturating NTP concentrations TSS-region sequence variation led to a ~100-fold range of relative expression (Figure 6B). Thus, the results show that limiting the concentrations of NTPs significantly enhances effects of TSS-region sequence variation on expression. Results further show that the magnitude of the difference in relative expression of TSS-region sequences carrying an R at TSS positions 7 and 8 relative to TSS-region sequence variants carrying a Y at TSS positions 7 and 8 increased at non-saturating NTP concentrations (Figure 6C). In addition, the magnitude of the difference in relative expression observed from TSS-region sequences carrying YR versus RY at positions 6/7 also increased at non-saturating NTP concentrations (Figure 6C). Furthermore, TSS-region sequences carrying 2–4 consecutive T bases, which exhibit high % slippage (Figure 5), exhibit a large decrease in relative expression at non-saturating NTP concentrations compared with saturating NTP concentrations (Figure 6C). Thus, we propose that this decrease in relative expression occurs, at least in part, due to a decrease in productive slippage synthesis and concomitant increase in non-productive slippage synthesis (undetectable by MASTER), as a consequence of performing reactions at non-saturating NTP concentrations. Our analysis of the effects of TSS-region sequence on transcript yield show that sequence variation in the TSS-region can impact overall transcript yields (i.e. promoter strength) by at least two orders of magnitude. We conclude that the TSS-region sequence is a key determinant of promoter strength.

MASTER analysis reveals a correlation between precision of TSS-selection and transcript yield

We next analyzed the relationship between TSS selection and transcript yields by comparing the mean TSS value with the relative expression value for each TSS-region sequence variant (Figure 7). This analysis revealed that sequence variants with the highest relative expression had mean TSS values extremely close to one of the two preferred TSS positions, position 7 and position 8 (Figure 7A–C, red arrows).

Next, we compared the variance from the mean TSS versus the relative expression for each sequence variant (Figure 7A–C, bottom). Results show that TSS-region sequences with the highest relative expression exhibited very low variance from the mean TSS (Figure 7, red circles; Figure S2F; $p < 0.001$). Thus, our MASTER analysis reveals that TSS-region sequences that exhibit the highest expression under a particular experimental condition also exhibit “precise” TSS selection at either position 7 or position 8. We conclude that precision of TSS selection is a previously undocumented contributor to promoter strength.

DISCUSSION

Here we report the development and application of MASTER, a technology leveraging the capabilities of high-throughput sequencing to enable the comprehensive analysis of the relationship between up to at least 4^7 (~16,000), and potentially at least 4^{10} (~1,000,000), DNA sequences and transcriptional output. MASTER provides many orders of magnitude higher throughput than conventional methods of interrogating the effects of DNA sequence on transcription output, can be carried out within a reasonable timescale, and can be performed at reasonable cost. We demonstrate the utility of MASTER by applying this method to the analysis of transcription initiation. Our results provide a complete description of the relationship between TSS-region sequence and TSS selection, slippage synthesis, and transcript yields for initiation at a consensus core promoter by *E. coli* RNAP *in vitro* and *in vivo*.

MASTER analysis provides a comprehensive description of transcription initiation

We documented three measurable outputs of transcription initiation: TSS position, productive slippage synthesis, and yields of full-length transcripts, for a library comprising 4^7 (~16,000) sequence variants of a consensus *E. coli* promoter. The results define full inventories of transcription start sites (“TSSomes”) of *E. coli* RNAP *in vitro* and *in vivo* (Tables S1–S7) and full inventories of transcripts generated by productive slippage synthesis (“slippomes”) of *E. coli* RNAP *in vitro* and *in vivo* (Tables S1–S7). Furthermore, our analyses of productive slippage synthesis provide, to our knowledge, the first comprehensive description of the extent of slippage *in vitro* and *in vivo* for any RNAP. Results indicate that slippage synthesis occurs from the majority of TSS-region DNA sequences and reveal slippage by increments of two nucleotides occurs at surprisingly high levels (Figures 5, S5, Tables S1–S7).

Our analysis of the effect of DNA topology on TSS-selection provides mechanistic insight into this process. In particular, single-molecule FRET experiments indicate that transcription-bubble size in the RNAP-promoter open complex (RPO) can vary (Robb et al., 2013). Based on the flexibility of RPO, it has been proposed that variability in the position of the TSS relative to core promoter elements is mediated by changes in the size of the unwound transcription bubble in RPO. Specifically, it has been proposed that TSS selection at downstream sites is mediated by transcription-bubble expansion (“scrunching”), and TSS selection at upstream sites is mediated by transcription-bubble contraction (“anti-scrunching”) (Robb et al., 2013). Our MASTER results indicate that negative supercoiling, which can provide a

driving force for transcription-bubble expansion, favors TSS selection at downstream positions (Figures 3, S4), consistent with the hypothesis that TSS-selection involves scrunching and anti-scrunching.

MASTER analysis of the effects of TSS-region sequences on yields of full-length transcripts show that TSS-region DNA sequences can have profound, up to 100-fold, effects on transcript yield (Figures 6, S6, Tables S1–S7). Furthermore the impact of TSS-region sequence on the range of expression varies in response to changes in NTP concentrations. Our findings that promoter TSS-region sequences can dictate a wide-range of expression levels suggests that these sequences serve as a reservoir of expression level diversity. Moreover, this diversity should easily be accessible to mutational processes and natural selection for tuning or altering promoter output.

MASTER analysis comparing TSS selection and transcript yield revealed a previously unknown relationship between precision of TSS selection and promoter strength (Figure 7). In particular, we found that sequences with the highest expression exhibit precise TSS selection at positions 7 or 8 (Figure 7, Tables S1–S7). Our findings indicate that optimal expression is provided, in part, by TSS-region sequences that promote efficient start site selection at a single position. The identification of the precision of TSS selection as a previously unidentified contributor to promoter strength exemplifies the utility of MASTER for illuminating aspects of transcription that would be difficult, if not impossible, to discover through conventional approaches.

MASTER as a new approach for comprehensive analysis of the relationship between nucleic-acid sequence and functional output

Analyses of the behavior of chromosomally encoded promoters using high-throughput approaches (e.g. RNA-seq, ChIP-seq) have provided a wealth of information regarding mechanisms employed by cells to regulate gene expression. However, every promoter is an evolved, unique sequence, which constrains the ability of researchers using such approaches to infer universal properties based on aggregate behavior. Furthermore, for sequence regions of more than a few bases, the total sequence diversity represented by all promoters in a genome is significantly less than the maximum theoretical diversity contained in an equivalent length of randomized DNA sequence. MASTER overcomes the inherent limitations imposed by analysis of chromosomally encoded promoters by enabling the comprehensive measurement of transcription output for all possible sequence variants of a given region of a transcription unit. In addition, MASTER allows the analysis of the behavior of all possible sequence variants in a region of interest to be performed over diverse conditions *in vitro* and *in vivo*.

Furthermore, although for sequence-specific promoter elements one can predict the effect of mutations on the overall affinity of RNAP for the promoter element, correlating the effect on transcription output is not straightforward. Thus, predicting how a given promoter will respond to alterations in conditions, and identifying the sequence determinants that dictate the response, represents an immense challenge. MASTER overcomes these limitations by enabling systematic variation of sequence attributes in a controlled fashion. Thus, we anticipate that results obtained from these and future studies using

MASTER will enable more accurate predictions of the behavior of chromosomally encoded promoters, and will inform the design of synthetic promoters for use in artificial biological circuits. In addition, MASTER can be readily adapted for comprehensive analysis of sequence determinants for transcription elongation, transcription termination, or any other biological process that involves complex nucleic-acid interactions.

Limitations

In this work, we have considered effects of TSS-region sequences on transcriptional output within the context of a consensus core promoter. However, in principle, the sequence of other promoter sequences (i.e. the UP-element, -35 element, the spacer region, the extended -10 element, the -10 element, and the discriminator element) also may impact transcription start site selection. The potential roles of other promoter sequences can be addressed simply by including additional regions of the promoter within the randomized segments. Currently, the method enables up to at least 7 bp, and potentially up to at least 10 bp, to be included in the randomized segment (Figures 1, S1). As sequencing technologies improve, the method will provide comprehensive coverage of even larger segments. In addition, the *in vivo* experiments reported in this work are performed in the context of a promoter that is plasmid-borne. Thus, a priority for future work will be to modify the MASTER procedure in order to analyze the behavior of promoter libraries *in vivo* in the context of the bacterial chromosome.

EXPERIMENTAL PROCEDURES

Construction of MASTER plasmid libraries

To generate library pMASTER-*lacCONS-N7* (Figures 1, S1A), a PCR product containing the *lacCONS* promoter with a 7-nt randomized region positioned 4–10 nt downstream of the -10 element and a 15-nt randomized “barcode” region positioned 27–41 nt downstream of the -10 element was subcloned into pSG289 (see Supplemental Experimental Procedures and Figure S1A). Sequencing analysis of the pMASTER-*lacCONS-N7* library indicated that it contains 16,295 (~99.5%) of a possible 16,384 sequences in the TSS-region (Figure S1C). Library pMASTER-*lacCONS-N10* was generated using a similar procedure and contains 1,019,505 (~97.2%) of a possible 1,048,576 sequences carrying a 10-bp randomized region (Figure S1B).

Generation of RNAs from the pMASTER-*lacCONS-N7* library *in vitro*

10 nM of template DNA (see Supplemental Experimental Procedures) was mixed with 50 nM RNAP holoenzyme in transcription buffer (50 mM Tris HCl pH 8.0, 10 mM MgCl₂, 0.01 mg/ml BSA, 100 mM KCl, 5% glycerol, 10 mM DTT, 0.4U/μl RNase OUT) and incubated at 37°C for 10 min. Transcription was initiated by adding NTPs (to a final concentration of 10 mM, 1 mM or 0.1 mM each NTP) and heparin (to 0.1 mg/ml). (We note that the Mg²⁺ concentration of 10 mM is limiting. Accordingly, the effective NTP: Mg²⁺ concentration in reactions performed in the presence of 10 mM NTPs is 2.5 mM for each NTP.) Reactions were stopped after 15 min by addition of EDTA to 10 mM. RNAs were analyzed by 5' RNA-seq as described in the Supplemental Experimental Procedures.

Generation of RNAs from the pMASTER-lacCONS-N7 library in vivo

DH10B-T1^R cells carrying pMASTER-*lacCONS-N7* were grown at 37°C in 50 ml of LB containing chloramphenicol (25 µg/ml) in a 250 ml flask shaken at 210 RPM. RNA was isolated at OD₆₀₀ ~0.5 and used for 5' RNA-seq analysis as described in the Supplemental Experimental Procedures. Plasmid DNA was also isolated from cells and used as a template in emulsion PCR reactions to generate a product that was sequenced to assign barcodes (see Supplemental Experimental Procedures).

High-throughput sequencing of RNA 5' ends (5' RNA-seq)

5' RNA-seq was done using procedures described in (Vvedenskaya et al., 2015) and in the Supplemental Experimental Procedures. cDNA libraries were sequenced using an Illumina HiSeq 2500.

Data analysis

Sequencing of template DNA was used to associate the 7-bp randomized sequence in the region of interest with a corresponding second 15-bp randomized sequence that serves as its barcode. For 5' RNA-seq analysis we considered only those reads that contained a perfect match to the sequence downstream of position 10 and a perfect match to the 5 bases downstream of the 15-bp barcode. The identity of the 15-bp barcode was used to determine the identity of bases at positions 4–10 of the *lacCONS* template. Sequences derived from the RNA 5' end of reads that were perfect matches to the sequence of the template were used for analysis of TSS selection. Sequences derived from RNA 5' ends that carried one or more mismatches from the DNA template and/or extended up to 5 bases upstream of position 4, were not considered in the analysis of TSS position, but were used for analysis of productive slippage synthesis and for analysis of transcript yields (for further details see Supplemental Experimental Procedures). All of the reads used for our data our analyses are provided in Tables S1–S7.

Source code and documentation for analysis of DNA templates and 5' RNA-seq libraries are provided at: <https://github.com/NickelsLabRutgers/MASTER-Data-Analysis>.

Data deposition—Raw reads have been deposited in the NIH/NCBI Sequence Read Archive under the study accession number SRP057850.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Craig Kaplan for discussion, Jeremy Bird and Maria Ciaramella for assistance with preparation of template libraries, Jared Knoblauch for preliminary data analysis, and Hanif Vahedian Movahed for RNAP. Work was supported by FIRB-Futuro in Ricerca RBF12001G_002 “Nematic” (AV and VV) and NIH grants GM041376 (RHE), GM088343 (BEN), GM096454 (BEN), and GM115910 (BEN).

REFERENCES

- Aoyama T, Takanami M. Essential structure of *E. coli* promoter II. Effect of the sequences around the RNA start point on promoter function. *Nucleic Acids Res.* 1985; 13:4085–4096. [PubMed: 2409530]

- Dangkulwanich M, Ishibashi T, Bintu L, Bustamante C. Molecular mechanisms of transcription through single-molecule experiments. *Chem. Rev.* 2014; 114:3203–3223. [PubMed: 24502198]
- Hawley DK, McClure WR. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.* 1983; 11:2237–2255. [PubMed: 6344016]
- Heyduk E, Heyduk T. Next generation sequencing-based parallel analysis of melting kinetics of 4096 variants of a bacterial promoter. *Biochemistry.* 2014; 53:282–292. [PubMed: 24359527]
- Heyduk T, Heyduk E. Next Generation Sequencing-based analysis of RNA polymerase functions. *Methods.* 2015
- Jacques JP, Kolakofsky D. Pseudo-templated transcription in prokaryotic and eukaryotic organisms. *Gene Dev.* 1991; 5:707–713. [PubMed: 2026325]
- Jeong W, Kang C. Start site selection at *lacUV5* promoter affected by the sequence context around the initiation sites. *Nucleic Acids Res.* 1994; 22:4667–4672. [PubMed: 7984416]
- Jorgensen SE, Buch LB, Nierlich DP. Nucleoside triphosphate termini from RNA synthesized *in vivo* by *Escherichia coli*. *Science.* 1969; 164:1067–1070. [PubMed: 4890175]
- Larson MH, Landick R, Block SM. Single-molecule studies of RNA polymerase. *Mol. Cell.* 2011; 41:249–262. [PubMed: 21292158]
- Lewis DE, Adhya S. Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. *Mol. Micro.* 2004; 54:692–701.
- Liu J, Turnbough CL. Effects of transcriptional start site sequence and position on nucleotide-sensitive selection of alternative start sites at the *pyrC* promoter in *Escherichia coli*. *J. Bacteriol.* 1994; 176:2938–2945. [PubMed: 7910603]
- Maitra U, Hurwitz H. The role of DNA in RNA synthesis, IX. Nucleoside triphosphate termini in RNA polymerase products. *Proc. Natl. Acad. Sci. U.S.A.* 1965; 54:815–822. [PubMed: 5324397]
- Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG Jr, Kinney JB, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* 2012; 30:271–277. [PubMed: 22371084]
- Murakami KS. Structural biology of bacterial RNA polymerase. *Biomolecules.* 2015; 5:848–864. [PubMed: 25970587]
- Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM, et al. Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nat. Biotechnol.* 2012; 30:265–270. [PubMed: 22371081]
- Robb NC, Cordes T, Hwang LC, Gryte K, Duchi D, Craggs TD, Santoso Y, Weiss S, Ebright RH, Kapanidis AN. The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. *J. Mol. Biol.* 2013; 425:875–885. [PubMed: 23274143]

- Ruff EF, Record MT Jr, Artsimovitch I. Initial events in bacterial transcription initiation. *Biomolecules*. 2015; 5:1035–1062. [PubMed: 26023916]
- Shultzaberger RK, Chen Z, Lewis KA, Schneider TD. Anatomy of *Escherichia coli* σ^{70} promoters. *Nucleic Acids Res*. 2007; 35:771–788. [PubMed: 17189297]
- Sorensen KI, Baker KE, Kelln RA, Neuhard J. Nucleotide pool-sensitive selection of the transcriptional start site in vivo at the *Salmonella typhimurium* *pyrC* and *pyrD* promoters. *J. Bacteriol*. 1993; 175:4137–4144. [PubMed: 8100568]
- Thomason MK, Bischler T, Eisenbart SK, Forstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM, Storz G. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J. Bacteriol*. 2015; 197:18–28. [PubMed: 25266388]
- Turnbough CL. Regulation of gene expression by reiterative transcription. *Curr. Opin. Microbiol*. 2011; 14:142–147. [PubMed: 21334966]
- Turnbough CL, Switzer RL. Regulation of pyrimidine biosynthetic gene expression in bacteria. *Microbiol. Mol. Biol. Rev*. 2008; 72:266–300. [PubMed: 18535147]
- Vvedenskaya IO, Goldman SR, Nickels BE. Preparation of cDNA libraries for high-throughput RNA sequencing analysis of RNA 5' ends. *Methods Mol. Biol*. 2015; 1276:211–228. [PubMed: 25665566]
- Walker KA, Osuna R. Factors affecting start site selection at the *Escherichia coli* *fis* promoter. *J. Bacteriol*. 2002; 184:4783–4791. [PubMed: 12169603]
- Wang F, Greene EC. Single-molecule studies of transcription: from one RNA polymerase at a time to the gene expression profile of a cell. *J. Mol. Biol*. 2011; 412:814–831. [PubMed: 21255583]
- Washburn RS, Gottesman ME. Regulation of transcription elongation and termination. *Biomolecules*. 2015; 5:1063–1078. [PubMed: 26035374]

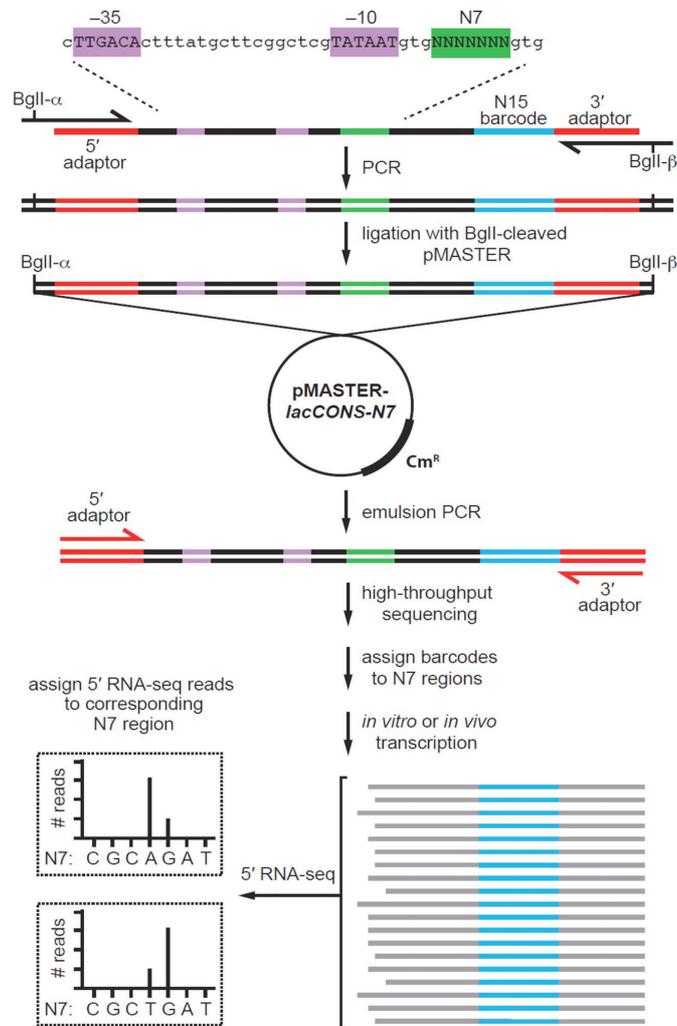


Figure 1. Massively systematic transcript end readout (MASTER) Top: generation of pMASTER-*lacCONS-N7* library. An oligodeoxyribonucleotide carrying the *lacCONS-N7* promoter and 15-nt barcode sequence (blue) is used as template in a PCR reaction using primers that introduce BglII sites. The BglII digested PCR product is cloned into BglII digested plasmid pSG289 (Figure S1A) to generate plasmid pMASTER-*lacCONS-N7*, which contains 4^7 (~16,000) sequences at positions 4–10 bps downstream of the *lacCONS* –10 element (green). Middle: product generated by emulsion PCR is used for high-throughput sequencing analysis to assign barcodes to TSS-sequence variants. PCR primers shown in red (5' and 3' adaptor) carry sequences that facilitate analysis using an Illumina HiSeq. Bottom: 5' RNA-seq analysis of RNA produced from the library *in vitro* and *in vivo*. The sequence of the barcode is used to assign the RNA to a TSS-region, the sequence of the 5' end is used to define the TSS, and the number of reads is used to measure transcript yield from each TSS-region sequence. (See Figure S1)

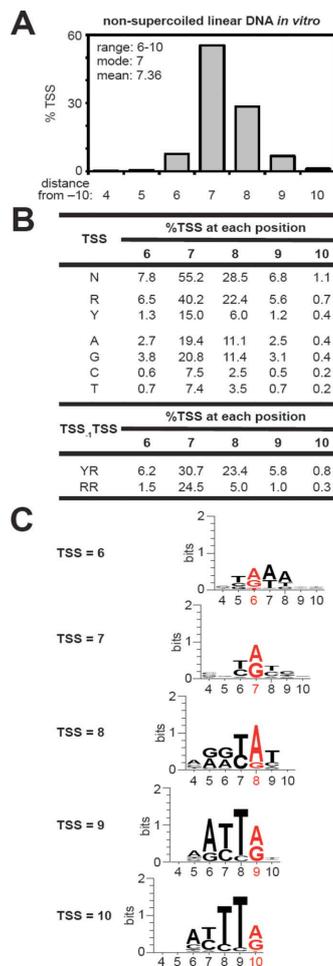


Figure 2. TSS selection on a non-supercoiled linear DNA template *in vitro* (A) TSS-distribution histogram. Average %TSS at positions 4–10 for TSS-regions with ≥ 25 matched RNA reads (Table S1). (B) Sequence determinants for TSS selection. Table lists the amount of the total %TSS at positions 6–10 derived from TSS-regions carrying (i) R or Y at the indicated TSS position, (ii) A, G, C, or T at the indicated TSS position, or (iii) YTSS-1RTSS or RTSS-1RTSS at the indicated TSS position with the highest %TSS at positions 6–10. Red bases indicate the TSS. (See Figure S3) (C) Sequence preferences for TSS selection. Sequence logo for the 162 TSS-region sequences (top 1%) with the highest % TSS at positions 6-10. Red bases indicate the TSS. See Figure S3.

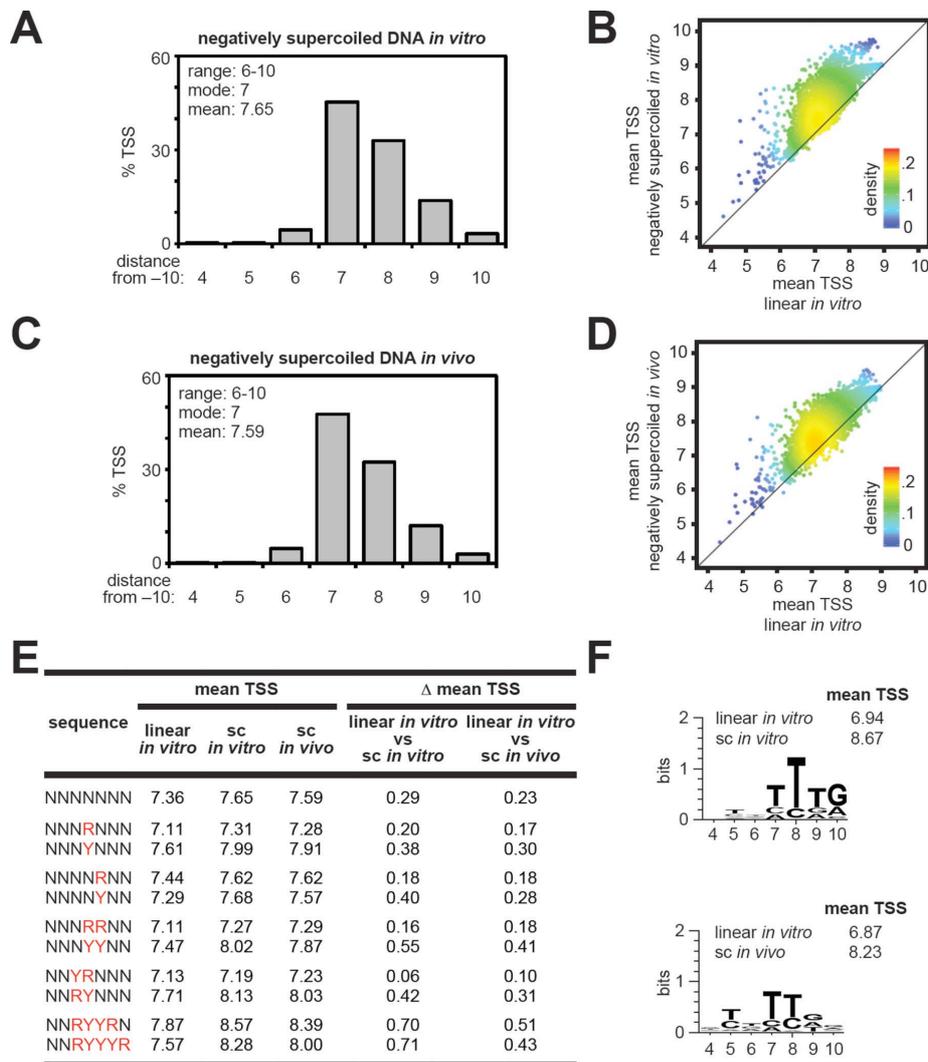


Figure 3. TSS selection on negatively supercoiled DNA templates (A) TSS-distribution histogram for experiments performed *in vitro*. Average %TSS at positions 4–10 for TSS-regions with ≥ 25 matched RNA reads (Table S2). (B) Plot of the mean TSS with negatively supercoiled DNA *in vitro* versus the mean TSS with non-supercoiled linear DNA *in vitro* for individual TSS-region sequences. (C) TSS-distribution histogram for experiments performed *in vivo*. Average %TSS at positions 4–10 for TSS-regions with ≥ 25 matched RNA reads (Table S3). (D) Plot of the mean TSS with negatively supercoiled DNA *in vivo* versus the mean TSS with non-supercoiled linear DNA *in vitro* for individual TSS-region sequences. (E) Average of the mean TSS values for the indicated TSS-region sequences. (Δ mean TSS; differences between values observed on linear and supercoiled templates). (F) Sequence preferences for topology dependent effects on TSS selection. Sequence logo and average mean TSS values for 162 TSS-region sequences (top 1%) with the highest values of Δ mean TSS. (See Figure S4).

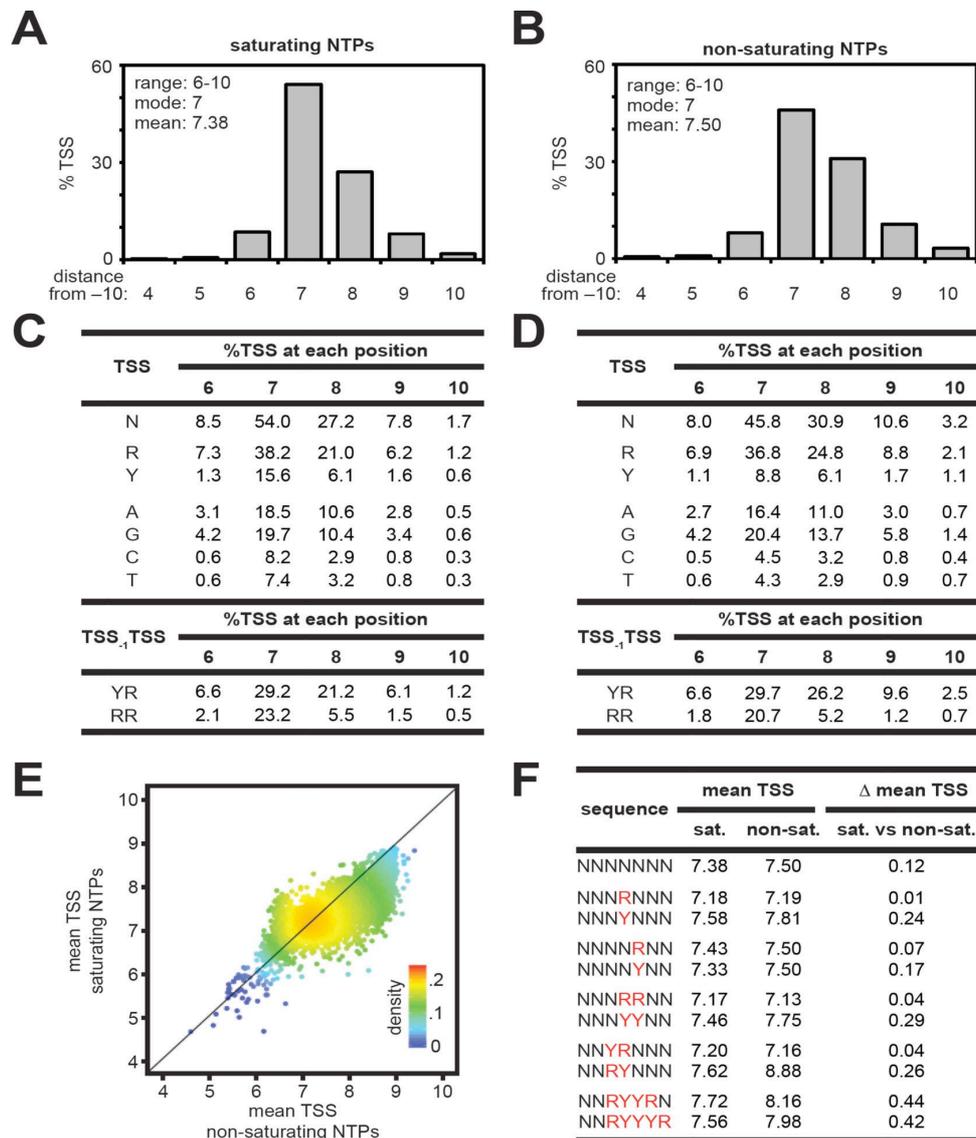


Figure 4. Effects of NTP concentrations on TSS selection *in vitro* (A. and B.) TSS-distribution histograms at saturating (A) and non-saturating (B) NTP concentrations *in vitro*. Average %TSS at positions 4–10 for TSS-regions with ≥ 25 matched RNA reads (Tables S4 and S5). Experiments were performed at 2.5 mM NTPs: Mg^{2+} (saturating) or 0.1 mM NTPs (non-saturating) using a non-supercoiled linear DNA template. (C. and D.) Sequence determinants for TSS selection. (C, saturating NTPs; D, non-saturating NTPs) (E) Plot of the mean TSS at saturating NTP concentrations versus non-saturating NTP concentrations for individual TSS-region sequences. (F) Average of the mean TSS values observed for the indicated TSS-region sequences at saturating (sat.) and non-saturating (non-sat.) NTP concentrations. (Δ mean TSS; differences between values observed at saturating and non-saturating NTP concentrations).

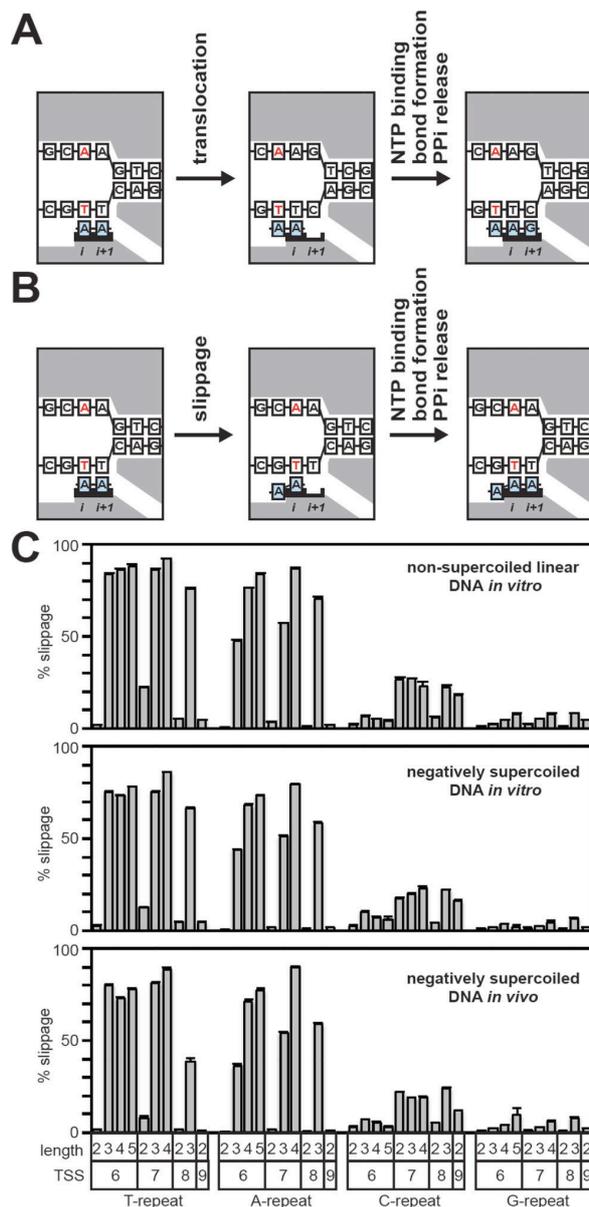


Figure 5. Comprehensive analysis of productive slippage synthesis. (A) Nucleotide addition cycle for the standard pathway of transcription initiation. Left: initial transcribing complex with a 2-nt RNA in a pre-translocated state. Middle: initial transcribing complex with a 2-nt RNA in a post-translocated state. Right: 3-nt product complex in a pre-translocated state. The RNA and DNA template strand remain in lock-step register and the sequence of the RNA is fully complementary to the template strand. White boxes, DNA; blue boxes, RNA; gray shading, RNAP; red, TSS bases; i and $i+1$, RNAP active-center i and $i+1$ sites. (B) Nucleotide addition cycle for the slippage pathway. Left: initial transcribing complex with a 2-nt RNA in a pre-translocated state. Middle: RNA has moved backward relative to the DNA template by one base. Right: 3-nt product complex in a pre-translocated state. The 5' end of the RNA carries an RNA/DNA difference and is not complementary to the template strand. (C) Analysis of productive slippage synthesis. Graphs show % slippage (mean + SEM) for TSS-region sequences containing 5' end homopolymeric repeat sequences of the indicated length that begin at the indicated position (TSS). (See Figure S5)

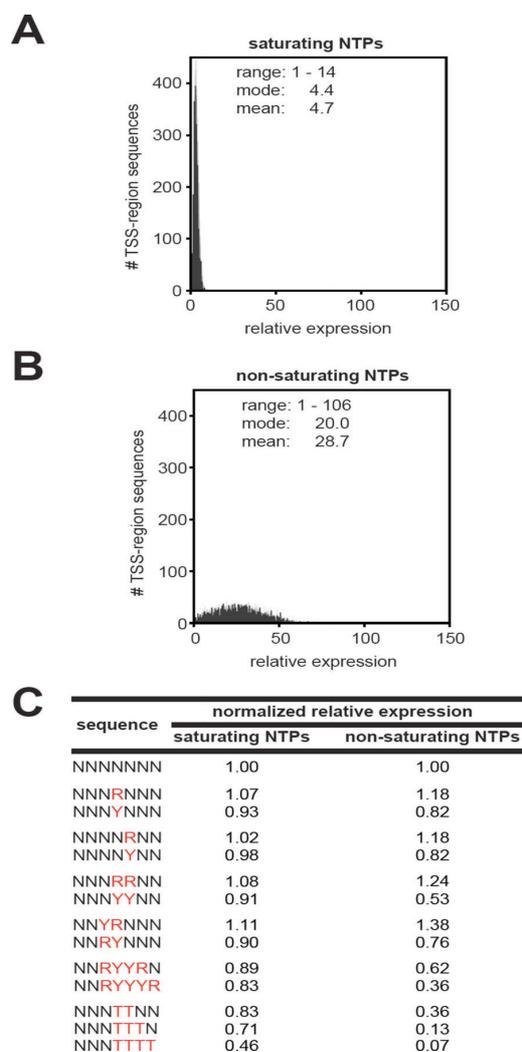


Figure 6. Effects of NTP concentrations on transcript yields *in vitro* (A. and B.) Relative expression histograms for experiments performed at saturating NTP (A) and non-saturating (B) NTP concentrations using a non-supercoiled linear DNA template *in vitro*. Relative expression for TSS-region sequences with ≥ 25 total RNA reads for which the number of DNA templates was not in the top or bottom 10% (Tables S4 and S5). For each experimental condition the lowest value of relative expression was normalized to 1. C. (C) Normalized relative expression for the indicated TSS-region sequences. Values were calculated by dividing the average relative expression for the indicated TSS-region sequence by the relative expression observed for all TSS-region sequences. (See Figure S6)

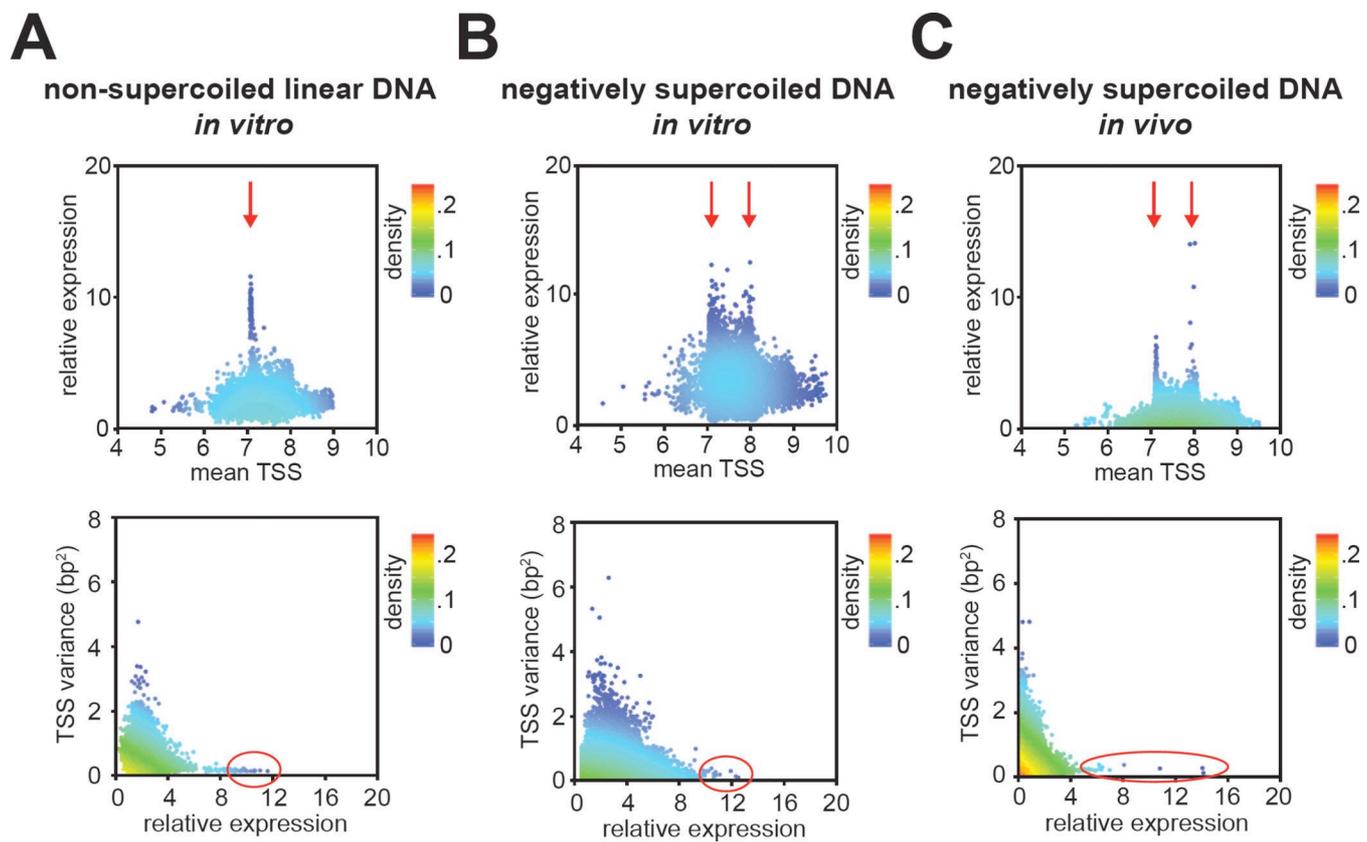


Figure 7. Precision of TSS selection is a determinant of transcript yield (A-C) Top: plot of relative expression versus mean TSS. Bottom: plot of TSS variance versus relative expression.

2.2 Comparison of Plasmid DNA Topology at nanoscale level

Visone V et al.

Manuscript

Abstract

DNA topology plays essential roles in several fundamental biological processes, such as DNA replication, recombination, and transcription. DNA topology describes the tertiary conformations of DNA that include supercoiling, knots, and catenanes. Despite its importance, however, much about supercoiled DNA is unknown. In mesophilic organisms, DNA is negatively supercoiled and, until recently, positive supercoiling was considered a peculiar mark of (hyper)thermophilic archaea needed to survive high temperatures. However, several lines of evidence suggest that negative and positive supercoiling might coexist in both (hyper)thermophilic and mesophilic organisms, raising the possibility that positive supercoiling might serve as a regulator of various cellular events. The atomic force microscope (AFM) is relatively new tool with enormous potential importance to study the biological structure and dynamic processes of DNA molecules. However, on the other hand, there is still no experimental data for positively supercoiled DNA molecules.

Here we apply for the first time the AFM to analyze positive supercoiled plasmid in comparison with other topologies. The AFM analysis in combination with biochemical approaches enabled us to perform an in deep characterization of DNA molecules with defined supercoiling degrees.

Introduction

The topological state of DNA and its impact on cellular activities has been recognized since the double helical structure of DNA was discovered (Watson and Crick 1953). The antiparallel configuration of two intertwined complementary strands provides genomic stability and protection, but it immediately raised questions regarding the topological complications associated with DNA metabolic processes (Yeonee Seol & Keir C. Neuman, 2016). DNA is organized into topologically closed domains or closed circular DNA such as plasmids (Dixon et al. 2012; Espéli and Boccard 2006) and as consequence its rotational motion is limited. Therefore, opening of the two DNA strands associated with many DNA metabolic processes generates downstream overwinding (positive supercoiling), and upstream underwinding (negative supercoiling), of the DNA. DNA topoisomerases provided a resolution to these topological problems (Champoux and Dulbecco 1972; Gellert et al. 1976; Wang 1979). By transiently breaking and resealing the DNA backbone via a transesterification reaction (Champoux 2001; Chen et al. 2013; Forterre et al. 2007; Schoeffler and Berger 2008), topoisomerases modulate the level of supercoiling and maintain the topological homeostasis of genomic DNA (Nitiss 2009; Vos et al. 2011; Wang 2002).

Supercoiled DNA is a closed circular double stranded DNA, where one strand usually winds around the other a specific number of times. Those topological properties are defined by few topological parameters: twist (T_w , the number of times each helix twist around the other) and writhe (W_r , the number of crossing

the double helix makes around itself). If DNA cannot break and resealed itself, the sum of these two parameters is a topological invariant, called linking number ($L_k = T_w + W_r$) (Witz and Stasiak 2010). Because the energy required to supercoil a polymer depends on its length, the density of supercoil is often used ($\sigma = (L_k - L_{k0}) / L_{k0}$, with L_{k0} is the L_k of the relaxed molecule). A positively supercoiled DNA has added (clockwise) twists, while a negatively supercoiled DNA has a reduced T_w (it has been counterclockwise twisted). The topological state of prokaryotic genomic DNA is strictly regulated to maintain negative supercoiling (Worcel and Burgi 1972). In eukaryotes, the supercoiling density is also effectively negative, but it is largely constrained within nucleosomes (Bates and Maxwell 1997; Travers and Muskhelishvili 2007). In *Escherichia coli*, σ in general ranges from -0.03 to -0.06 depending on the growth phase (Bliska and Cozzarelli 1987; Hatfield and Benham 2002). The existence of alternative topological states was demonstrated by the discovery of reverse gyrase, a DNA topoisomerase inducing positive supercoiling and the finding of episomal DNA in relaxed or positively supercoiled form in several hyperthermophilic archaea (Lulchev and Klostermeier 2014). These findings suggested that, in each organism, the actual DNA topology is the result of the adaptive pressure to maintain the proper equilibrium between genome dynamics and stability under the physicochemical conditions specific to that particular organism. However, recent results suggest that negative and positive supercoiling might coexist in both (hyper)thermophiles as well as mesophiles (Valenti et al. 2011). In general it is thought that underwound DNA facilitates the strand separation required for DNA transactions, while DNA overwinding affects the helical repeat, helix melting and/or flexibility, resulting in DNA stabilization (Valenti et al. 2011). Actually, the mechanism of positive supercoiling has not been completely established, however it is known that the rate and extent of positive supercoiling are limited by the positive torque on the DNA (Ogawa et al. 2015, 2016). Thus, the knowledge of how DNA supercoil is of great importance both to understand biological mechanisms and to support computational models of DNA and to confirm they are predictive.

A number of experiments have been performed by pulling DNA bond to magnetic beads to understand how the molecule accommodates this extra torsional stress (Meng et al. 2014; Strick et al. 1998; Strick et al. 1996). Unfortunately the results of this type of experiments do not directly relate with the topological state of DNA plasmids, because the force required to stress the molecule is an external stimulus that does not exist in plasmids physiological conditions. Thus the way plasmids supercoil and which is the topological state that minimizes their energy is still a matter of debate. More recently electron cryo-tomography microscopy have been used to investigate small minicircles (Irobalieva et al. 2015). They notice that the minicircles show a broad distribution of conformations and positive and negative supercoiling are accommodated differently. The cryo-ET, together with biochemical, and computational studies demonstrated the astounding versatility and dynamism of DNA depending on the degree of supercoiling. These results suggest that the different conformation of DNA could have a key role during

DNA metabolic processes. However there is still no data for supercoiled DNA at the larger scales (several of kilo base pairs) that could be more representative of DNA found in nature.

Plasmid DNA is broadly used as a model in DNA research because it has three different conformations: supercoiled forms, relaxed forms, and linear forms. Gel electrophoresis is considered one of the best methods to separate these plasmids because different topological structures have different radii of gyration and electrophoretic mobilities, allowing them to separate by running on a gel. Recent research showed that these conformations can also be identified using AFM imaging and the ratio of different forms can be counted directly at a single molecule level (Jiang et al. 2010). The atomic force microscope (AFM) was invented by Binnig et al. in 1986. After that, its quick development resulted in the commercial AFM available to the biological and chemical researchers. In the last decades, AFM has proven to be one of most valuable techniques for the studies of DNA and the researchers were mostly focused on the biological structure and dynamic processes of DNA. By analyzing the topological structure of each DNA molecule from AFM images, the distributions of DNA with different configurations can be identified directly and easily. Due to this advantage, AFM imaging can also be ideal tool for quantifying DNA structure under a wide range of DNA amounts, concentrations and sizes that may not able to be handled by other methods. Thus, atomic force microscopy is a viable alternative to gel electrophoresis in the study of different DNA structures.

Although a large amount of data is present regarding the AFM analysis of negatively supercoiled plasmids (Lyubchenko et al. 1997), data for positive supercoiling are not available. Using a combination of biochemical approaches we unveil a strong asymmetry in the conformation of negatively and positively supercoiled DNA, which can be highlighted with AFM, using proper deposition procedures. While gel electrophoresis is considered the *de-facto* method to separate plasmids, AFM characterization reaches higher level of detail and it describes fine topological differences. Biological implications of these findings are discussed.

Results

Determination of DNA superhelical densities

The superhelical density (σ) permits direct comparison of DNA molecules differing in size. Moreover it allows estimating the topological state of the plasmids under physiological conditions. In our work, the σ values of the plasmids with different topology were estimated by the band counting method after one-dimensional agarose gel electrophoresis using partially relaxed DNA as marker. pBluescript DNA plasmid (pBs) containing ~3000 bp was selected for these studies.

The negatively supercoiled pBs was extracted from *E. coli* cell culture growth at 37°C. The sigma value was experimentally determined by electrophoresis gel in presence of chloroquine using the following equation: $\sigma = \Delta L_k / L_{k0}$ (see below).

The positively supercoiled pBs was obtained by incubating the negative pBs with *Pyrobaculum calidifontis* reverse gyrase topoisomerase. In order to determine the superhelical density of positively supercoiled DNA, one-dimensional agarose gel electrophoresis with no drug was performed. We deduced a σ value of -0.039 for the negatively supercoiled DNA and a σ value of $+0.05$ for the positively supercoiled DNA (Figure 1).

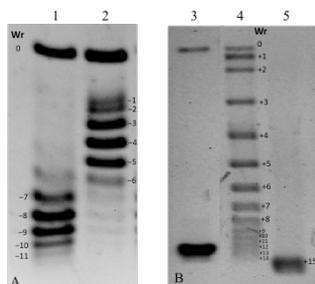


Fig.1: gel electrophoresis to determine superhelical density of negatively and positively supercoiled pBs plasmid. Gel A) in presence of $1 \mu\text{g/mL}$ chloroquine; gel B) without intercalating agents. Lane: 1) negatively supercoiled pBs, 2) partially negative supercoiled pBs, 3) negatively supercoiled pBs, 4) partially positive supercoiled pBs, 5) positively supercoiled pBs.

We failed to produce more positively or negatively supercoiled plasmids under a broad range of conditions suggesting that obtained supercoiling is representative of the maximum supercoiled DNA produced by enzymatic reaction *in vitro* (Figure 2).

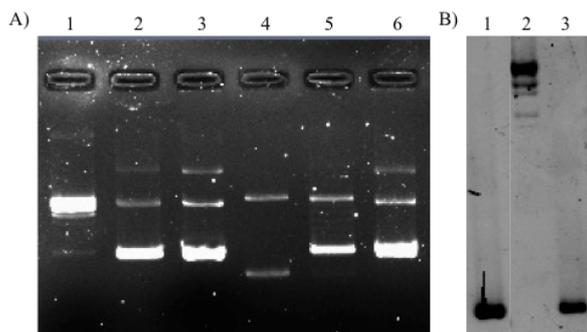


Fig.2: A) DNA gyrase assay starting from relaxed or positively supercoiled pBs. Lane: 1) relaxed pBs, 2) relaxed pBs incubated with DNA gyrase, 3) negatively supercoiled pBs extracted from *E. coli* (control), 4) positively supercoiled pBs, 5) positively supercoiled pBs incubated with DNA gyrase, 6) negatively supercoiled pBs extracted from *E. coli* (control). B) Reverse gyrase (RG) assay starting from relaxed pBs. Lane 1) negatively supercoiled pBs incubated with RG (control), 2) relaxed pBs, 3) relaxed pBs incubated with RG.

Biochemical properties of positively and negatively supercoiled plasmid

In order to characterize the biochemical properties of DNA molecules with opposite supercoiling degree we set up biochemical assays using specific enzyme or chemical molecule. For the first the stability of plasmids with different topology was evaluated by assessing the presence of exposed bases. Bal31 is a known single-strand DNA exonuclease, capable of endonuclease activity on double-strand DNA containing helical disruptions. Mainly, Bal31 recognizes single-stranded components of melted duplex DNA therefore providing a powerful tool for detecting helix-destabilized DNA. The negatively supercoiled DNA was incubated with Bal31 nuclease at 30°C for different time. After 30 minutes a large amount of plasmid was degraded and after 1 hour completely disappeared (Figure 3). Consistent with its Bal31 sensitivity, the electrophoretic mobility of the negatively supercoiled DNA shifted when incubated with glyoxal, a small molecule that traps exposed bases. At 1 M concentration of glyoxal all plasmid was shifted (Figure 4). Taken together our results demonstrate that negatively supercoiled plasmid contains single strand regions in line with its biological role. What about the positively supercoiled one? When treated with Bal31, positively supercoiled plasmid was not degraded up to 60 minutes of incubation suggesting that the plasmid doesn't contain single strand regions (Figure 3). Furthermore the electrophoretic mobility of the positively supercoiled plasmid was not shifted when incubated with increasing concentrations of glyoxal suggesting that the overwound DNA inhibits strand separation (Figure 4).

These results demonstrate that in our conditions the positive supercoiling is more stable than negative protecting DNA from nuclease degradation and from helix breathing. This data supports the potential role of positive supercoiling in DNA stability maintenance at high temperature.

In order to evaluate the effect of temperature on positive supercoiling we set up a glyoxal experiment at different temperatures. Both negatively and positively supercoiled plasmids were incubated for 5' minutes at increasing temperatures (from 80°C to 100°C) in the presence of glyoxal. For the underwound DNA the glyoxal shift increases with temperature. Surprisingly the overwound DNA was not bound up to 100°C suggesting that the positive supercoiling protects DNA from thermal denaturation and could be a key role *in vivo* in thermophilic organisms (Figure 5).

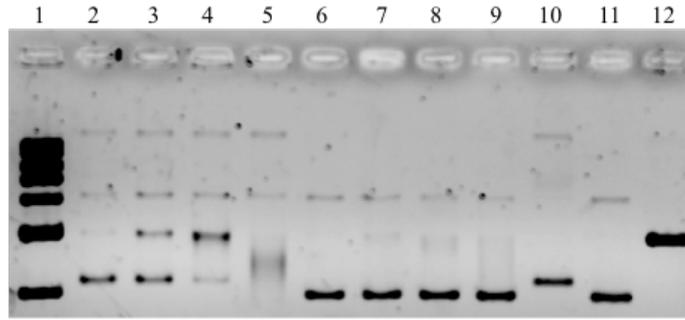


Fig.3: Nuclease Bal-31 assay. Lane 1) DNA Ladder, from 2) to 5) negatively supercoiled pBs incubated for increasing time (1, 10, 30, 60 minutes) with Bal-31, from 6) to 9) positively supercoiled pBs incubated for increasing time (1, 10, 30, 60 minutes) with Bal-31, 10) untreated negatively supercoiled pBs, 11) untreated positively supercoiled pBs, 12) pBs linearized with EcoRI.

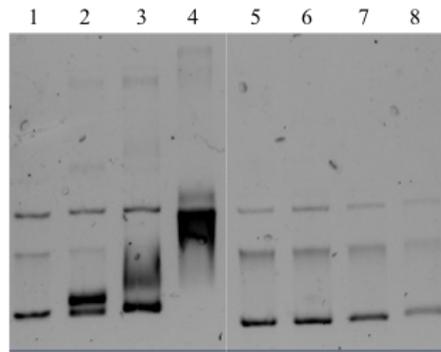


Fig.4: Glyoxal binding assay. Lane 1) and 5) negatively and positively supercoiled pBs control; from 2) to 4) negatively supercoiled pBs incubated with increasing concentrations of glyoxal (0.25, 0.5, 1 M), from 5) to 8) positively supercoiled pBs incubate incubated with increasing concentrations of glyoxal (0.25, 0.5, 1 M).

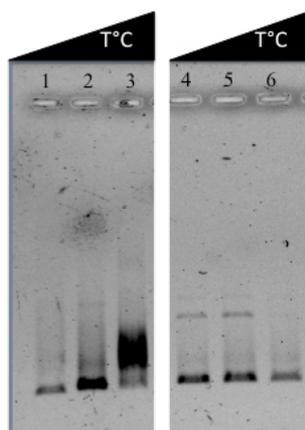


Fig.5: Glyoxal binding assay at high temperature. Lane 1, 2 and 3 negatively supercoiled pBs incubated with 1 M glyoxal for 5 minutes at 80, 90 and 100°C respectively; Lane 4, 5 and 6 positively supercoiled pBs incubated with 1 M glyoxal at 80, 90 and 100°C respectively.

Physical properties of positively and negatively supercoiled plasmid: AFM analysis

Several lines of evidence suggest that the atomically smooth surface and stable chemical properties of Mica mineral make it an ideal substrate for DNA adsorption, and it has been commonly used in experimental research, such as studies of DNA conformation by AFM analysis. DNA binding to mica has been shown to be enhanced in the presence of millimolar concentrations of divalent metal counterions (Lyubchenko 2011).

In order to study the DNA plasmids with different conformations by AFM analysis, the DNA molecules were deposited on mica substrate in presence of $MgCl_2$. As showed in Figure 6, in our conditions the two types of molecules assume significantly distinct shapes showing a more compact conformations when positively supercoiled respect to the more relaxed conformation of negatively supercoiled molecules.

In order to quantitatively analyze the different conformations we developed a routine that traces the plasmids and extracts a set of geometrical parameters useful to classify the different type of supercoiling. We considered two parameters able to distinguish molecules with different supercoiling degree: the "density" which represents the ratio between weighted area and filled area of each plasmid and the "Euler number" which specifies the number of objects in the region minus the number of holes in those objects. As reported in Figure 7 the analysis of both parameters showed that the two types of supercoiled molecules are statistically different.

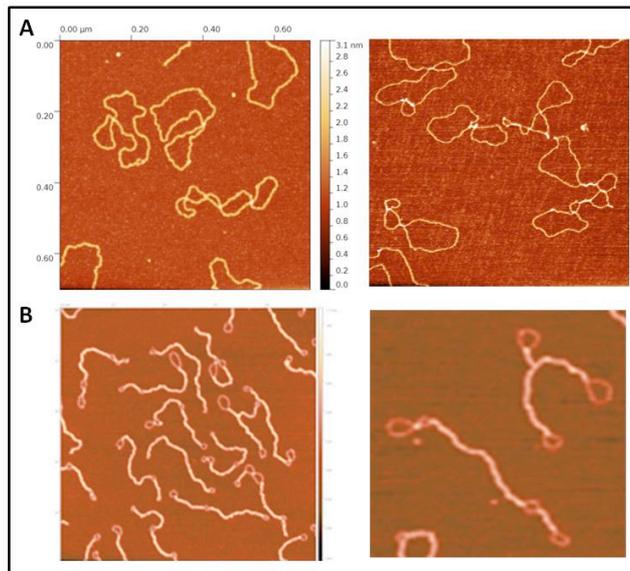


Fig.6: zoomed AFM images of supercoiled pBluescript DNA (3 kb). The images were collected on air-dried samples. (A) Negatively supercoiled plasmids directly extracted from *E. coli*. (B) Positively supercoiled plasmids obtained by reaction with reverse gyrase. All the samples were deposited on mica in presence of 3 mM $MgCl_2$ and successively dried.

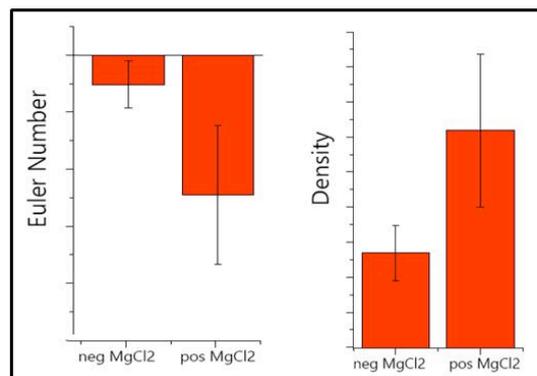


Fig.7: quantitative parameters obtained from the analyses of AFM images of almost 100 plasmids both negatively and positively supercoiled. The differences between the two topologies and the relative error bars are shown.

From relaxed to positive supercoiled: the reverse gyrase reaction

In order to analyze the conformational transitions from negative to positive supercoiled plasmids we performed a time course analysis of reverse gyrase reaction. The negatively supercoiled plasmid was incubated in presence of *Sulfolobus solfataricus* reverse gyrase (SsRG) at 85°C for different times. After reaction the topoisomerase activity was analyzed by two-dimensional agarose gel electrophoresis, which allows separation of positive and negative plasmid topoisomers (Figure 8). Such assay allows qualitative and quantitative analysis of topoisomerase activity, taking into account both the amount of plasmid

topological isomers (topoisomers) produced and their linking number (L_k), and gives information not only on the efficiency, but also on the distributive or processive character of the reaction.

As showed in Figure 8 SsRG activity shows a distributive behavior introducing positive supercoils progressively with increasing time of reaction. At 5 minutes almost all DNA is converted in positively supercoiled molecule. In order to make an in-deep analysis of positive supercoiling reaction, we determined the "writhe number" (number of supercoils) of the most abundant topoisomer for each reaction time (Figure 9). The quantitative analysis of SsRG activity showed that number of supercoils increase gradually with time, supporting the distributive behavior proposed for reverse gyrase.

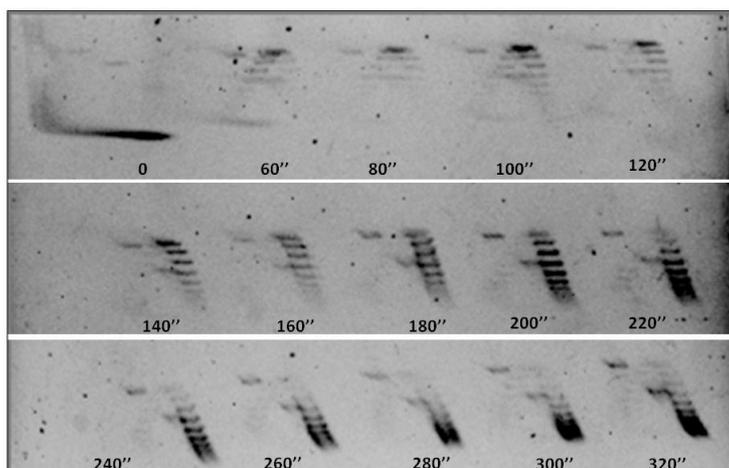


Fig.8: Two-dimensional gel of SsRG activity after indicating time of reaction.

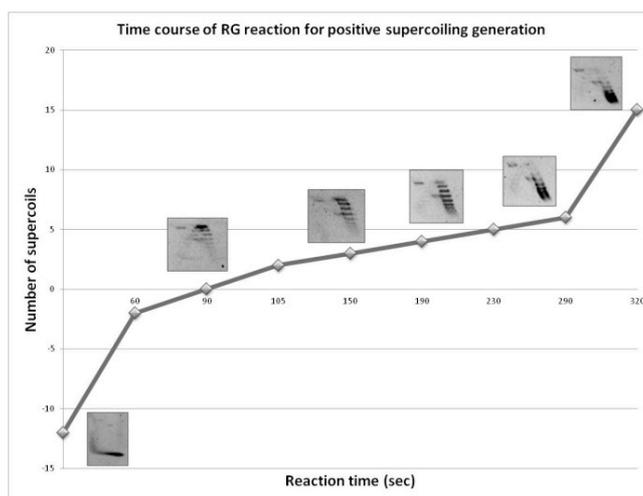


Fig. 9: Quantitative analysis of SsRG reaction. The number of supercoils (writhe) is showed in function of the reaction time (seconds). The average value of supercoils for each population of topoisomers was calculated with the method of the most abundant topoisomer within each population. A gel imaging of the selected reaction times is showed.

The supercoiled plasmids produced by SsRG were then analyzed by atomic force microscopy. The Figure 10 shows an AFM sample images of negatively supercoiled plasmid before (panel A) and after incubation with SsRG at indicated times (panel B-E). The images show a time dependent topology modification from negative to positive plasmid highlighting the transitions among the different conformations.

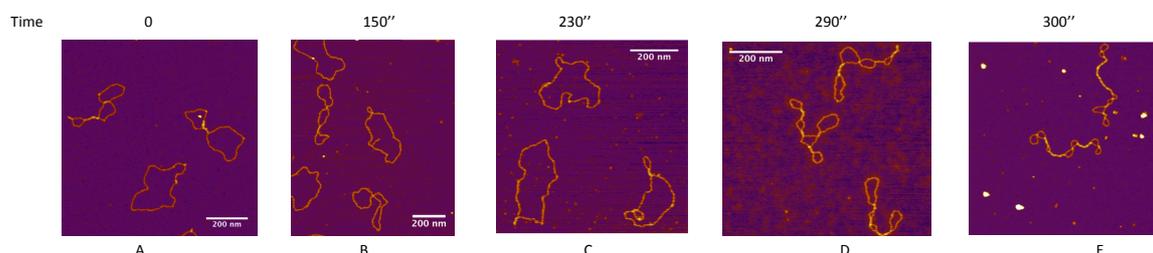


Fig.10: AFM images of negative supercoiled pBluescript DNA (3 kb) before (panel A) and after (B-E) incubation with SsRG at indicated reaction times. The topological modifications from negative to positive are showed. At 300 sec a higher positively supercoiled degree is obtained.

AFM analysis of plasmids in liquid

One of the most important features of AFM is the ability to scan samples in aqueous solutions. This mode of AFM imaging is appealing for two major reasons. First, imaging of the fully hydrated sample eliminates potential problems inherent with sample drying. Second, imaging in liquid allows for time-lapse visualization of sample dynamics and DNA-protein interaction (Lyubchenko et al. 2014). To improve our understanding about the stability and dynamics of DNA molecules with specific conformations, we set up the experimental conditions to analyze the DNA structures by AFM in liquid medium, under physiological conditions.

We reported in Figure 11 for the first time AFM images of positively supercoiled DNA plasmid performed in solution. As showed, both negatively and positively supercoiled plasmids appear more flexible; however, while negatively supercoiled molecules assume a wider range of conformations, positively supercoiled molecules tend to adopt a “chain-like” structure (Figure 11).

Importantly, these results are in line with the data obtained in air experiment suggesting that the different conformations of positively and negatively supercoiled are independent of deposition method. The ability to visualized both positive and negative topology under physiological conditions represent a starting point for further studies about proteins interaction with specific DNA topologies.

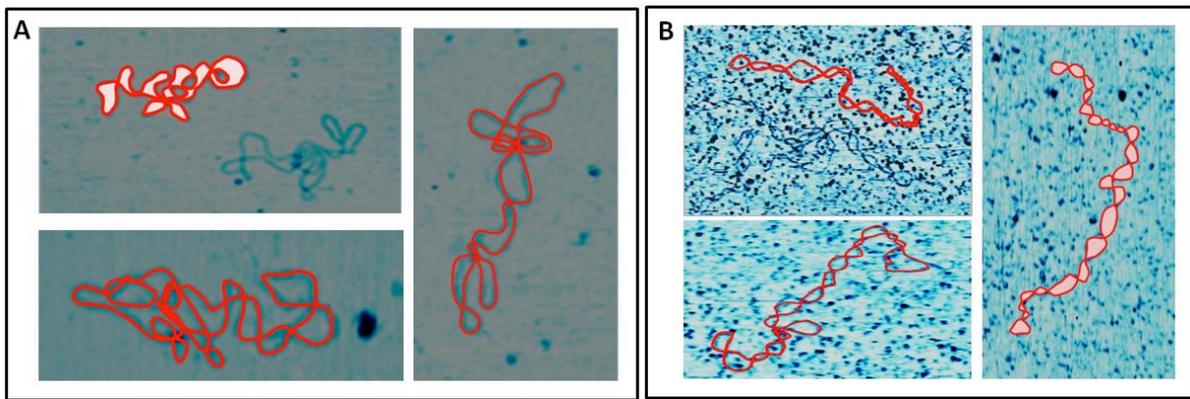


Fig.11: Liquid and zoomed AFM images of supercoiled pBluescript DNA (3 kb). The images were collected in buffer. (A) Negatively supercoiled plasmids directly extracted from *E. coli*. (B) Positively supercoiled plasmids obtained by reaction with reverse gyrase.

Discussion

Since its discovery, supercoiled DNA has been extensively studied by biologists and by physicists. The former were interested in its implication in biological processes, while the latter investigated its structure and studied its topology and conformation. DNA supercoiling plays key roles in genetic processes in the cell such as replication, transcription, recombination as well as DNA packaging (Sinden 1994). A lot of studies was performed in order to clarify the structural properties of supercoiled DNA and large amounts of experimental data are available in literature. In the last decade, Atomic Force Microscopy (AFM) has proven to be one of most valuable techniques for the studies of the biological structure and dynamic processes of DNA. Despite its great potential, all studies performed until now used as substrate linear or negatively supercoiled DNA molecules while no studies on positively supercoiled DNA molecules have been performed by AFM. In my work I analyzed for the first time the positively supercoiled DNA in comparison with the negatively by AFM analysis in order to investigate structural peculiarities of both topologies. Moreover I analyzed both conformations also by classical biochemical techniques.

For the first the biochemical analyses revealed the presence of exposed bases when the DNA was negatively supercoiled but not in positively supercoiled DNA. As already know, the exposed bases create a local helix distortions that allow DNA to assume conformations that would be energetically unfavorable if the bases remained paired. The presence of single strand regions in negative supercoiled DNA is in line with its biological role during replication and transcription. Conversely the absence of exposed bases in the positive supercoiled molecules suggests that the positive conformation could play a key role in DNA stability maintenance preventing the local denaturation and protecting the DNA structure.

AFM data showed different conformations for negatively and positively DNA in almost all the conditions used. In particular I observed that in presence of a low concentration of divalent cations (3 mM $MgCl_2$), negatively supercoiled DNA molecules (with a superhelical density of -0.04) tend to adopt

loosely interwound conformations (as already reported in Schmatko et al. 2014) while positively supercoiled DNA (with a superhelical density of +0.05) showed a typical plectonemic conformation. It could be hypothesized that the different electrostatic repulsion at crossings causes a different immobilization of supercoiled molecules on mica substrate: in the case of negatively supercoiled plasmids the repulsion exceeds bending and twisting costs. Therefore, a DNA molecule would prefer to be under high twist rather than having crossings. In the case of positively supercoiled, the electrostatic repulsion at crossings is probably reduced because of the different orientation of supercoiling and as a consequence we observe plectonemes (Figure 12).

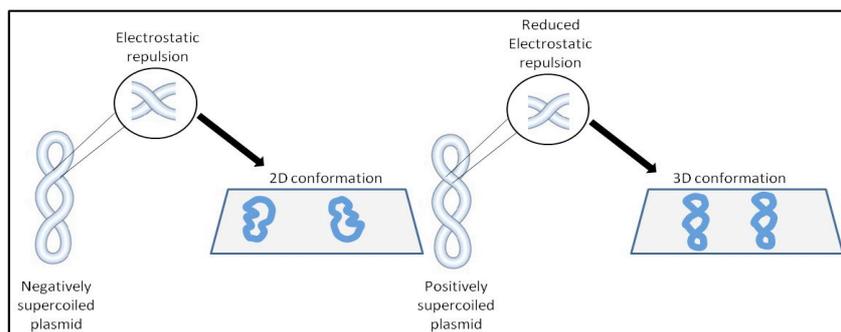


Fig.12: hypothetical model to explain the different behavior of negatively and positively supercoiled DNA when immobilized on mica substrate in presence of a low concentration of divalent cations (3 mM MgCl₂). *Left:* the electrostatic repulsion at crossings forced negatively supercoiled plasmid to assume a more open conformation (2D conformation). *Right:* in positively supercoiled plasmid the presumable reduce electrostatic repulsion enables it to preserve a plectonemic shape (3D conformation).

Taken together our results demonstrate that positively and negatively supercoiled plasmid assume significantly distinct shapes showing more compact conformations when positively supercoiled respect to the more relaxed of negative form. These properties support the potential role of positive supercoiling in hyperthermophilic organisms protecting DNA against thermal denaturation. Moreover the compact conformation could have a general role during DNA transactions preventing the local opening and modulating the protein binding.

In conclusion the obtained results represent a starting point for further studies to identify possible differences between negative and positive supercoiling and to elucidate their biological roles *in vivo*.

Materials and Methods

Determination of DNA superhelical densities

DNA superhelical densities were estimated by the band counting method using partially relaxed DNA as marker. The negatively supercoiled pBs was extracted from *E. coli* cell culture growth at 37°C using Invitrogen Midi prep kit. The sigma value was experimentally determined by electrophoresis gel in

presence of 1 µg/mL chloroquine. Sc⁺ plasmid was obtained by incubating purified negative supercoiled plasmid (Sc⁻) with *Pyrobaculum calidifontis* Reverse gyrase topoisomerase. The incubation was for 10' at 90°C at protein/DNA molar ratio of 3 in presence of ATP 1mM (Jamroze et al., 2014). After incubation, Sc⁺ plasmid was purified by phenol extraction and ethanol precipitation and checked by analysing an aliquot of the sample by two-dimensional gel electrophoresis in 1.2% agarose and 1X TBE (Napoli et al., 2005). Partially positive supercoiled forms were prepared using reverse gyrase activity at different incubation times. In order to obtain relaxed plasmid, purified Sc⁺ plasmid was incubated with wg-Topoisomerase I (Promega) at protein/DNA ratio of 20U/µg for 30' at 37 °C in 50mM Tris-HCl (pH 7.5), 50mM NaCl, 0.1mM EDTA, 1mM DTT. After incubation, the reaction was stopped with 2% of SDS and relaxed plasmid DNA was purified by phenol extraction and ethanol precipitation. The extent of relaxed plasmid was checked by analysing an aliquot of the sample by mono-dimensional gel electrophoresis in 1.2% agarose and 1X TBE. Gel was stained with ethidium bromide (1 µg/ml) and destained in deionized water for the analyses with UV light by a Bio-Rad VersaDoc apparatus. In order to determine the superhelical density of positively supercoiled DNA, one-dimensional agarose gel electrophoresis with no drug was runned.

Nuclease Bal-31 assay

Negatively and positively supercoiled DNA (400 ng in 40 ul final volume) were incubated with 0.2 units of nuclease Bal-31 at 30 °C in 20mM Tris-HCl, pH 8.0, 600mM NaCl, 12mM MgCl₂, 12mM CaCl₂ and 1mM disodium EDTA. At 1, 10, 30 and 60-minute intervals, 10 ul (100 ng) samples were removed, quenched by addition of an equal volume of stop buffer (50mM Tris-HCl, pH 8.0, 100mM disodium EDTA, 10% glycerol, 200 µg/mL proteinase K), followed by incubation at 45 °C for 30 min to degrade Bal-31. Products were analysed by electrophoresis through 1.2% agarose in 1X Tris-Borate EDTA. Gel was stained with ethidium bromide (1 µg/ml) and destained in deionized water for the analyses with UV light by a Bio-Rad VersaDoc apparatus.

Glyoxal assay

Glyoxal was first deionized with AG-501-X8 mixed bed ion-exchange resin (Bio-Rad, Hercules, CA). 100 ng of negatively or positively supercoiled DNA was incubated with increasing concentrations of glyoxal (from 0,2 M to 1M) in 10mM sodium phosphate, pH 7.0, for 16 h at room temperature. Control reactions were incubated in 10mM sodium phosphate only. Samples were analysed by electrophoresis through 0,8% agarose gel in 1X Tris-Borate EDTA. Gel was stained with ethidium bromide (1 µg/ml) and destained in deionized water for the analyses with UV light by a Bio-Rad VersaDoc apparatus.

DNA Gyrase assay

DNA gyrase assay was performed using relaxed or positively supercoiled plasmid pBluescript as substrate. Briefly, in a total volume of 20 µL, 1 U of DNA Gyrase (Sigma) was incubated with 300 ng of the substrate in 1x DNA gyrase buffer (35 mM Tris-HCl pH 7.5, 24 mM KCl, 20 mM MgCl₂, 1 mM ATP, 2 mM DTT, 1,8 mM Spermidine, 6.5% glycerol, 100 µg/ml BSA) at 37°C for ½ h. Samples were

analysed by one-dimensional agarose gel electrophoresis. After electrophoresis, gels were stained with ethidium bromide (1 mg/ml), analyzed and quantified under UV light with a Chemi-doc apparatus and the QuantityOne software (Bio-Rad, Hercules, CA, USA).

Positive supercoiling assay

Positive supercoiling assay was performed using plasmid pBluescript (Qiagen) as substrate. Briefly, in a total volume of 20 μ L, 60 nM of protein was incubated with 300 ng of the substrate (negatively supercoiled or relaxed) in 1x reverse gyrase buffer (35 mM Tris-HCl, pH 7.0, 0.1 mM Na₂EDTA, 30 mM MgCl₂, 2.0 mM DTT, 1 mM ATP) at 85°C for 10 min. Samples were analysed by one-dimensional agarose gel electrophoresis. After electrophoresis, gels were stained with ethidium bromide (1 mg/ml), analyzed and quantified under UV light with a Chemi-doc apparatus and the QuantityOne software (Bio-Rad, Hercules, CA, USA). For time course experiment, we scale-up the conditions of reaction to a final volume of 300 μ L and we done withdrawals of 20 μ L every 10 seconds after the first minute of reaction. Samples were then analysed by two-dimensional agarose gel electrophoresis. Briefly, we performed the first dimension in absence of intercalating agent. The second dimension was perpendicular to the first and was runned in presence of ethidium bromide (0,02 mg/mL). After electrophoresis, gels were stained with ethidium bromide (1 mg/ml), analyzed as described above.

AFM Sample Preparation

Negatively supercoiled pBluescript was extracted E. coli culture through Invitrogen Midi-prep kit and further purified by ethanol precipitation. Positively supercoiled pBluescript was obtained by incubation with Pyrobaculum calidifontis reverse gyrase in 1x reverse gyrase buffer (see above) at 90°C for 10 minutes and then purified by phenol treatment and ethanol precipitation. The DNA preparations were mixed with ultrapure water to achieve a final concentration of 5 ng/ μ L in 10 μ L volume. The solution was then deposited on mica surface in presence of 3 mM MgCl₂ for 5 min. Afterward the mica was rinsed with 1 mL of ultrapure water and dried under a gentle nitrogen flow.

AFM Imaging

AFM images were collected using a MultiMode SPM with a Nanoscope III controller (Veeco Instruments, Santa Barbara, CA, USA) operated in tapping-mode in air. The AFM cantilevers had a spring constant of 5 N/m (Bruker cantilevers, TAP150A) with resonance frequencies ranging between 120 and 160 kHz. Each experiment was performed at least in triplicates and AFM images were obtained at several separate mica locations.

AFM in Liquid

The samples for the liquid experiments were prepared following the very same preparation protocol as used in the air experiments. Before the measurements, we immersed the samples in a working buffer (1 mM Tris HCl, 3 mM MgCl₂, pH 7.0). This ensured a loose attachment of the DNA to the mica surface, thus a stable AFM imaging of the molecules, while allowing their dynamic rearrangement.

References

1. Bates AD, Maxwell A. DNA topology: topoisomerases keep it simple. *Curr Biol.* 1997 Dec 1;7(12):R778-81. Review.
2. Bliska JB, Cozzarelli NR. Use of site-specific recombination as a probe of DNA structure and metabolism in vivo. *J Mol Biol.* 1987 Mar 20;194(2):205-18.
3. Champoux JJ, Dulbecco R. An activity from mammalian cells that untwists superhelical DNA--a possible swivel for DNA replication (polyoma-ethidium bromide-mouse-embryo cells-dye binding assay). *Proc Natl Acad Sci U S A.* 1972 Jan;69(1):143-6.
4. Champoux JJ. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem.* 2001;70:369-413. Review.
5. Chen SH, Chan NL, Hsieh TS. New mechanistic and functional insights into DNA topoisomerases. *Annu Rev Biochem.* 2013;82:139-70. doi: 10.1146/annurev-biochem-061809-100002. Review.
6. Depew DE, Wang JC. Conformational fluctuations of DNA helix. *Proc Natl Acad Sci U S A.* 1975 Nov;72(11):4275-9.
7. Espéli O, Boccard F. Organization of the Escherichia coli chromosome into macrodomains and its possible functional implications. *J Struct Biol.* 2006 Nov;156(2):304-10. Epub 2006 Aug 2.
8. Forterre P, Gribaldo S, Gadelle D, Serre MC. Origin and evolution of DNA topoisomerases. *Biochimie.* 2007 Apr;89(4):427-46. Review.
9. Gellert M, Mizuuchi K, O'Dea MH, Nash HA. DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc Natl Acad Sci U S A.* 1976 Nov;73(11):3872-6.
10. Hatfield GW, Benham CJ. DNA topology-mediated control of global gene expression in Escherichia coli. *Annu Rev Genet.* 2002;36:175-203.
11. Irobalieva R N, Fogg J M, Catanese Jr D J, Sutthibutpong T, Chen M, Barker A K, Ludtke S J, Harris S A, Schmid M F, Chiu W, Zechiedrich L. Structural diversity of supercoiled DNA. *Nature Comm.*, 6, 8440 (2015).
12. Jiang Y, Rabbi M, Mieczkowski P A, Marszalek P E. Separating DNA with Different Topologies by Atomic Force Microscopy in Comparison with Gel Electrophoresis. *J. Phys. Chem.*, 114, 12162-12165(2010).
13. Liu LF, Wang JC. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A.* 1987 Oct;84(20):7024-7.
14. Lulchev P, Klostermeier D. Reverse gyrase--recent advances and current mechanistic understanding of positive DNA supercoiling. *Nucleic Acids Res.* 2014 Jul;42(13):8200-13. doi: 10.1093/nar/gku589.
15. Lyubchenko Y L, Shlyakhtenko L S. Visualization of supercoiled DNA with atomic force microscopy *in situ*. *Proc. Nat. Acad. Sci.*, 94, 496-501(1997).

16. Lyubchenko YL, Gall AA, Shlyakhtenko LS. Visualization of DNA and protein-DNA complexes with atomic force microscopy. *Methods Mol Biol.* 2014;1117:367-84. doi: 10.1007/978-1-62703-776-1_17.
17. Meng H, Bosman J, Heijden T and Noort J. Coexistence of Twisted, Plectonemic, and Melted DNA in Small Topological Domains. *Biophys. J.*, 106, 1174-1181 (2014).
18. Nitiss JL. DNA topoisomerase II and its growing repertoire of biological functions. *Nat Rev Cancer.* 2009 May;9(5):327-37. doi: 10.1038/nrc2608. Review.
19. Ogawa T, Sutoh K, Kikuchi A, Kinoshita K Jr. Torsional stress in DNA limits collaboration among reverse gyrase molecules. *FEBS J.* 2016 Apr;283(8):1372-84. doi: 10.1111/febs.13675.
20. Ogawa T, Yogo K, Furuike S, Sutoh K, Kikuchi A, Kinoshita K Jr. Direct observation of DNA overwinding by reverse gyrase. *Proc Natl Acad Sci U S A.* 2015 Jun 16;112(24):7495-500. doi: 10.1073/pnas.1422203112.
21. Schmatko T, Muller P, Maaloum M. Surface charge effects on the 2D conformation of supercoiled DNA. *Soft Matter*,10, 2520-2529 (2014).
22. Schoeffler AJ, Berger JM. DNA topoisomerases: harnessing and constraining energy to govern chromosome topology. *Q Rev Biophys.* 2008 Feb;41(1):41-101. doi: 10.1017/S003358350800468X. Review.
23. Seol Y, Neuman KC. The Dynamic Interplay Between DNA Topoisomerases and DNA Topology. *Biophys Rev.* 2016 Sep;8(3):221-231.
24. Strick T R, Allemand J F, Bensimon D, Croquette V. Behavior of Supercoiled DNA. *Biophys. J.*, 74, 2016-2028 (1998).
25. Strick T R, Allemand J F, Bensimon D, Croquette V. The elasticity of a single supercoiled DNA molecule. *Science*, 271, 1835-1837 (1996).
26. Travers A, Muskhelishvili G. A common topology for bacterial and eukaryotic transcription initiation? *EMBO Rep.* 2007 Feb;8(2):147-51. Review.
27. Valenti A, Perugini G, Rossi M, Ciaramella M. Positive supercoiling in thermophiles and mesophiles: of the good and evil. *Biochem Soc Trans.* 2011 Jan;39(1):58-63. doi: 10.1042/BST0390058. Review.
28. Vos SM, Tretter EM, Schmidt BH, Berger JM. All tangled up: how cells direct, manage and exploit topoisomerase function. *Nat Rev Mol Cell Biol.* 2011 Nov 23;12(12):827-41. doi: 10.1038/nrm3228. Review.
29. Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol.* 2002 Jun;3(6):430-40. Review.
30. Wang JC. DNA: bihelical structure, supercoiling, and relaxation. *Cold Spring Harb Symp Quant Biol.* 1979;43 Pt 1:29-33.
31. Wang JC. Helical repeat of DNA in solution. *Proc Natl Acad Sci U S A.* 1979 Jan;76(1):200-3.

32. Witz G, Stasiak A. DNA supercoiling and its role in DNA decatenation and unknotting. *Nucleic Acids Res.*,38(7), 2119-2133(2009).
33. Worcel A, Burgi E. On the structure of the folded chromosome of *Escherichia coli*. *J Mol Biol.* 1972 Nov 14;71(2):127-47.

2.3 Heterologous expression of thermophilic reverse gyrase in an archaeal model system by exploiting a new thermostable protein tag

Visone V. et al.

Manuscript

Abstract

Reverse gyrase (RG) is a peculiar thermophilic DNA topoisomerase with the unique capability to introduce positive supercoils in DNA molecules. Although a more refined picture of the mechanism of positive supercoiling has emerged from the extensive studies in the past years, the reverse gyrase mystery has not entirely been solved. The *in vitro* and *in vivo* detection of proteins is fundamental to study their function. In the last decade, SNAP-tag™ technology was used as a valid and powerful biotechnological tool for the *in vivo* and *in vitro* specific labeling of proteins. This was made possible by the irreversible reaction of the alkylguanine-DNA-alkyl-transferase (hAGT) in the presence of benzyl-guanine derivatives. However the general instability of the mesophilic SNAP-tag™ make this new approach not fully applicable to extremophilic organisms. It was previously demonstrated that the engineered variant of the thermostable alkylguanine-DNA-alkyl-transferase shows a catalytic efficiency comparable to the SNAP-tag™ protein, but displays a higher intrinsic stability at elevated temperatures. Moreover the successful heterologous expression obtained in a thermophilic model organism (the bacterial *Thermus thermophilus*) makes SsOGT-H⁵ tag a valid candidate for organisms living in extreme conditions. Here for the first time, we use the archaeal *Sulfolobus islandicus* as model system for the heterologous expression of *Sulfolobus solfataricus* RG fused to SsOGT-H⁵ tag. This finding opens the possibility to an in-deep study of RG biological role *in vivo*.

Introduction

Reverse gyrase is a DNA topoisomerase with the unique capability to introduce positive supercoils into DNA in an ATP-dependent reaction. It consists of a helicase domain and a topoisomerase domain that closely cooperate in catalysis. All reverse gyrases isolated and characterized to date catalyze positive supercoiling of DNA *in vitro*. Although this reaction is considered the hallmark reaction of reverse gyrase, it is unclear whether positive DNA supercoiling occurs and what is its *in vivo* function (Lulchev and Klostermeier 2014). A lot of studies also support a role of reverse gyrase in DNA protection and repair. Indeed it was reported that reverse gyrase is degraded after treatment of *S. solfataricus* with alkylating agent, parallel to the degradation of genomic DNA (Valenti et al. 2006); furthermore it interacts with and inhibits a translesion polymerase in *S. solfataricus* (Valenti et al. 2009) and is recruited to DNA after ultraviolet irradiation (Napoli et al. 2004). However the most recent studies are focused on the mechanical features of positive supercoiling reaction. In particular, Ogawa et al. (2015, 2016) observed the overwinding reaction under a microscope by tethering a magnetic bead to a coverslip surface with a linear DNA and pulling the bead upward with a magnet. In the presence of reverse gyrase and ATP, the bead rotated continuously to relax the DNA wound by reverse gyrase. The rotation was

slow at 71 °C and was not observed at 60 °C or below. The introduction of 30 consecutive mismatched base pairs in the DNA, stimulates the overwinding reaction, as already demonstrated by gel assay, and the reaction persisted down to 50 °C. By changing the reverse gyrase concentration, they showed that the reverse gyrase binds the bubble for more than 5 min and when rotation is prohibited the bead sank as the bubble DNA was positively supercoiled by reverse gyrase. These results indicated that the overwinding rate is highly dependent on the tension and torsional stress in the DNA. Although these studies offer a mechanistic insight in positive supercoiling activity much remains to be understood about biological role of reverse gyrase. The *in vivo* detection of reverse gyrase may help uncover the physiological function of reverse gyrase and the role of its hallmark reaction in the physiological context.

Labeling proteins with synthetic probes is an important approach to study protein function (Gronemeyer et al. 2006). One way to label proteins is through expression of the protein of interest as a fusion protein with an additional polypeptide (called “tag”) that mediates the labeling (Johnsson and Johnsson, 2003). Previously human O⁶-alkylguanine-DNA alkyltransferase (hAGT) have been used as a tag for the specific labeling of proteins in living cells as well as *in vitro* (Keppler et al., 2003). The labeling occurs by the reaction of O⁶-benzylguanine (BG) derivatives with AGT; this reaction leads to the irreversible transfer of the derivatized benzyl group to the active site of AGT (Fig.1).

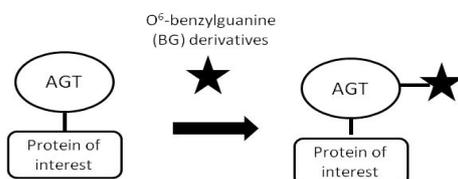


Fig.1: General scheme for labeling of AGT fusion proteins; adapted from Gronemeyer et al. 2006.

In my laboratory was previously characterized an AGT from the thermoacidophilic archaeon *Sulfolobus solfataricus* (*SsOGT*), which has the same catalytic efficiency of hAGT with fluorescent BG-derivatives, but displays a marked stability over a wide range of conditions (Perugino et al. 2012). Recently Vettone et al. use an engineered variant of the thermostable *SsOGT* protein, which lacks DNA binding activity (*SsOGT-H⁵*), as a protein-tag for extremophilic organisms. This discovery opened the possibility to widen the SNAP-tag™ technology to organisms, which live in extreme environments. Vettone et al. reported that *SsOGT-H⁵* is correctly folded, expressed, functionally active and stable in both mesophilic *E. coli* and thermophilic *T. thermophilus* hosts. Furthermore their results demonstrated that it is possible to study both mesophilic and thermophilic proteins/enzymes fused to *SsOGT-H⁵* under their own physiological conditions, without the need to remove the tag. These properties, together with the complete abolition of binding to DNA, make *SsOGT-H⁵* a robust alternative as protein-tag for the application of the SNAP-tag™ technology for *in vivo* and *in vitro* studies.

Here we show the heterologous expression of *Sulfolobus solfataricus* RG (Ss-RG) protein fused to thermostable SNAP-tag and its *in vivo* analysis in the thermophilic archaeon *Sulfolobus islandicus*.

Results

Ss-RG fusion with SsOGT-H⁵

In order to obtain a reverse gyrase fused with H⁵ thermophilic tag, *Sulfolobus solfataricus top1* gene was fused downstream to and in frame with the *SsOgt-H5* gene, for the expression of the H⁵-SsRG fusion protein in the *E. coli* ABE C strain. The protein expression was detected by western blot analyses (data not shown). Purification of the fusion protein by affinity chromatography through a nickel column was performed, exploiting the presence of a His6-tag at the N-terminus of the H⁵ moiety. Interestingly the presence of SsRG does not interfere with the H⁵ activity, making it possible to follow all purification steps by SDS-PAGE and fluorescence imaging of the samples, after incubation with the BG-FL substrate (Fig.2).

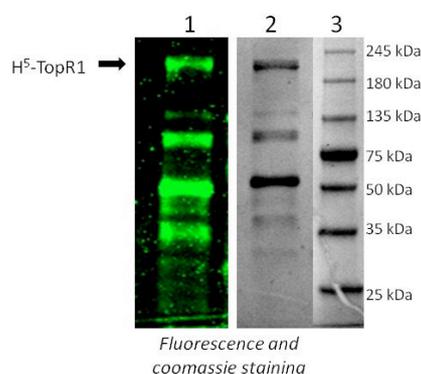


Fig.2: *E. coli* ABE C strain was transformed by using pQE-*ogtH5-topr1* plasmid. SDS-PAGE of the expression and purification by His6-tag affinity chromatography of the H⁵-SsRG fusion protein: pool of fractions eluted by imidazole was incubated with BG-FL (5 μ M). After, the gel was exposed for the fluorescence imaging analysis (1) and then stained by coomassie blue (2), as described in “Materials and Methods”. (3) Protein molecular weight marker.

Furthermore the utilization of the thermostable protein-tag enabled us to proceed with a thermoprecipitation treatment to purify the pool of eluted fractions and to perform analysis of RG enzymatic activity at high temperatures without the need to remove the tag.

The ability of fusion protein to introduce positive supercoils into DNA was checked incubating the purified protein in standard conditions as previously report for RG alone. As showed in Fig.3 the fusion protein exhibits the same activity of RG alone indicating that the tag does not influence the positive supercoiling activity of the topoisomerase.

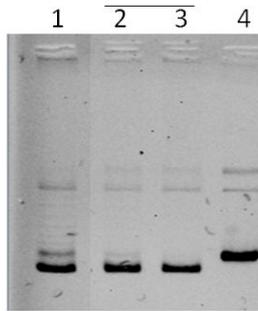


Fig.3: Positive supercoiling activity assay: lane (1) negatively supercoiled DNA incubated with purified reverse gyrase for 10 min. at 85°C; lane (2), (3) negatively supercoiled DNA incubated with the H5-SsRG 5 and 10 μ L, respectively; lane (4) negatively supercoiled DNA.

Heterologous expression of *SsOgt-H5* and *in vivo* labeling in *S. islandicus*

In order to evaluate the heterologous expression of SsOgt-H5 in *S. islandicus*, the gene was cloned in the pSeSD plasmid (Fig.4) containing: i) both *S. islandicus* and *E. coli* replication origins; ii) the amp gene for growth in ampicillin selective medium; iii) a multi-cloning site (MCS) for the expression of the SsOGT-H⁵ protein in *S. islandicus ogt* (-) strain and iv) *pyrEF* marker for the selection of transformants.

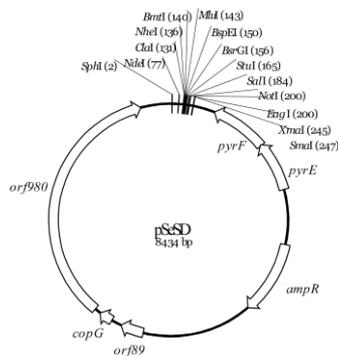


Fig.4: pSeSD plasmid used to transform *S. islandicus* strain.

In order to avoid a cross activity of endogenous *ogt*, we used, for the plasmid expression, a *S. islandicus* mutant strain in which the *ogt* gene was deleted (Δ -OGT). We set up the experimental conditions to transform the plasmid into Δ -OGT strain by electroporation. The obtained cells containing the plasmid were grown in minimal medium (to allow the selection of transformants) at 75°C and recovered after 5 days at \sim 0.7/0.8 OD.

As showed before for bacterial cells (Perugino et al. 2012; Vettone et al. 2016), it was possible to check the *in vivo* activity of H⁵, by directly incubating intact cells in the SCV medium (0.2% sucrose, 0.2% casamino acids plus a vitamin mixture) supplemented with the BG-FL substrate. To test the permeability of *S. islandicus* to this substrate, we incubated the cells at different temperatures and also in the presence of non-toxic organic solvents, such as DMSO, to improve cell permeabilization. In Figure 5 the SDS-PAGE of *S. islandicus* cells after incubation with BG-FL substrate is showed. The gel reveals a fluorescent band of the expected molecular weight in the cells from *S. islandicus/pSeSD-ogtH5*

transformants, whereas no signals were seen in cells transformed with the empty plasmid (pSeSD), confirming the lack of any endogenous alkyltransferase activity. The maximal activity was obtained at 70°C, likely due to the higher permeability of the cells at their physiological temperatures concomitant with an increase of activity of the thermophilic H⁵ protein (Fig.5).

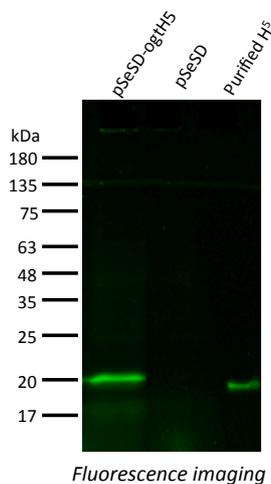


Fig.5: *S. islandicus* cells were transformed with the empty vector pSeSD and with the pSeSD-*ogtH5*. The protein expression was checked directly in the cells after incubation with BG-FL (5µM) by monitoring the fluorescence signal. After SDS PAGE the gel the was exposed for the fluorescence imaging analysis showing the H⁵ signal only in the cells transformed with pSeSD-*ogtH5*.

The fluorescent labeling of ogt-H⁵ could be seen also in intact *S. islandicus* cells by fluorescent microscopy analysis (Fig.6A). *S. islandicus ogt (-)* mutant transformed with the empty vector was used as control (Fig.6B).

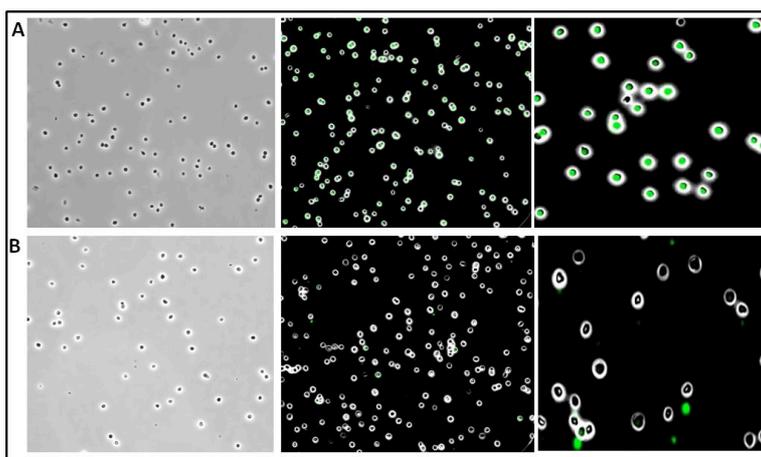


Fig.6: Fluorescent microscopy of *Sulfolobus islandicus* cells incubating with BG-FL (3µM). (A) *S. islandicus ogt (-)* mutant transformed with pSeSD-*ogtH5* plasmid (B) *S. islandicus ogt (-)* mutant transformed with the empty vector. Images show DIC and AlexaFluor488 (green).

As showed in Fig. 6, no signals were seen in the cells transformed with the empty plasmid, confirming the lack of any endogenous alkyltransferase activity and the absence of fluorescence background. Conversely a strong fluorescent signal appears in the cells containing the *ogtH⁵* gene indicating that the labeling is specific for H⁵ and that the protein is correctly folded, expressed, and functionally active in living cells.

H⁵-SsRG detection in living cells of *S. islandicus*

In order to detect reverse gyrase *in vivo*, H⁵-SsRG coding sequence was cloned into pSeSD plasmid and the resultant vector was used to transform *Sulfolobus islandicus ogt (-) pyrEF (-)* mutant. The transformants were grown at 75°C until stationary phase (OD ~ 0.8) and then incubated in the growth medium supplemented with the BG-FL substrate for the analyses at fluorescent microscopy (Fig.7) as described before.

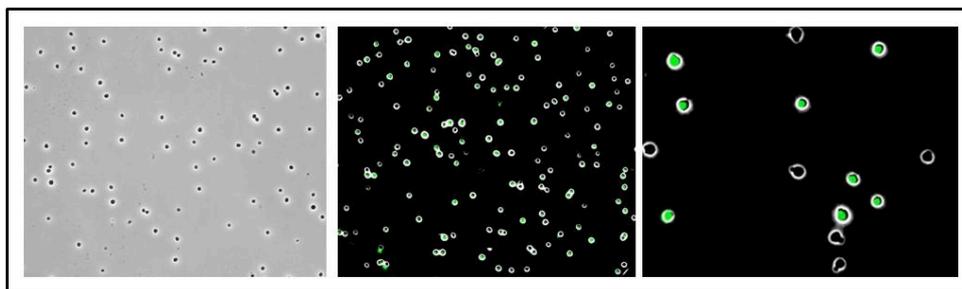


Fig.7: Fluorescent microscopy of *Sulfolobus islandicus ogt (-)* cells transformed with pSeSD-*ogtH⁵-Sstopr1* plasmid incubating with BG-FL (3µM). Images show DIC and AlexaFluor488 (green).

The images show for the first time the successful heterologous expression of RG gene in *S. islandicus*. Moreover the presence of H⁵-tag opens the possibility to an in-deep analysis of RG expression and localization in physiological conditions in living cells.

Discussion

Labeling proteins is a fundamental approach to study proteins function. Before the discovery of SNAP-tag technology, the green fluorescent protein (GFP) and its derivatives have been extensively applied in biological field (Chalfie et al. 1994; Tsien 1998). However GFP has various limitations such as its relative big dimensions, which might affect the target protein's function and localization. Furthermore the isolation of the natural fluorophore in the active site makes GFPs and variants partially sensitive to environmental changes; because the formation of the internal fluorophore is oxygen-dependent, GFPs are not suitable tags in applications requiring anaerobic conditions and finally their general is restricted to mesophiles and mild reaction conditions.

Mechanistic studies on reverse gyrase have focused on the catalysis of positive DNA supercoiling by reverse gyrase *in vitro*. However to our knowledge the reverse gyrase has never been seen *in vivo*.

Here we demonstrated the heterologous expression of RG protein fused to the thermostable SNAP-tag. The analysis of the H⁵-RG fusion protein demonstrated that it is possible to study the RG under its own physiological conditions, without the need to remove the tag. This new fusion protein gives the opportunity for further *in vivo* and *in vitro* studies that could elucidate all the possible biological roles of reverse gyrase (positive supercoiling, repair, protection).

The application of SNAP-tagTM technology will be useful for RG analysis *in vivo* by protein-protein interaction, in situ localization, FRET experiment in combination with thermostable GFP variants, etc. and also for *in vitro* studies about protein-protein interaction and purification.

Materials and Methods

Reagents

All chemicals were purchased from Sigma-Aldrich; SNAP-Vista GreenTM substrate (referred to BG-FL) was from New England Biolabs (Ipswich, MA). Synthetic oligonucleotides were from Primm (Milan, Italy) and listed in Table 1, *Pfu* DNA polymerase were from Stratagene (La Jolla, CA), as well as *Escherichia coli* ABLE C strain.

DNA constructs

In order to insert *ogtH5* gene in pSeSD vector, *ogtH5* gene was PCR-amplified from the pQE-*ogtH5* construct (Perugino et al. 2012) using H5-up NdeI and H5-dw Sall primers (Table 1) which possess NdeI and Sall sites, respectively. This allowed us to insert *ogtH5* gene in the multi-cloning site of pSeSD plasmid for the heterologous expression of H⁵ in *S. islandicus*. For the fusion protein H⁵-RG, the DNA fragment of the *ogtH5* gene was PCR-amplified from the pQE-*ogtH5* construct (Perugino et al. 2012) using H5-fwd and H5-rev oligonucleotides (Table 1), which possess a *SacI* site. This allowed the ligation of this gene upstream and in frame with the *topr1* gene in the pQE-*topr1* construct (Valenti et al. 2008), and leading to the pQE-*ogtH5-topr1* plasmid for the heterologous expression of the SsOGT-H5-SsRG fusion protein in *E. coli*. pSeSD-H⁵-*SstopR1* for the heterologous expression in *S. islandicus* of the H⁵-SsRG fusion protein was obtained by multiple rounds of PCR amplification: the pQE_upstream-fwd/H5_Nde-rev oligonucleotides pairs (Table 1), were first used to introduce *NdeI* site upstream H5 sequence. In a second round of PCR, the H5_Nde-fwd/H5-rev oligonucleotides pairs (Table 1), were used to obtain a DNA fragment that overlaps the product of the first PCR round. Finally the former two DNA fragments were fused to each other by using the external pQE_upstream-fwd/H5-rev oligonucleotides pair. Finally, the obtained DNA fragment was ligated in the multi-cloning site of the pSeSD vector by using the *NdeI/NheI* sites. For all the obtained constructs, regions encoding the cloned genes were verified by DNA sequencing (Primm, Milan, Italy).

Primers	Sequence
H5-up NdeI	5'-GCGATATCCATATGCTGGTGTATGGATTGTATAAAAAG-3'
H5-dw Sall	5'-GTACGTCGACTTCTGGAATTTTGACTCCTTCC-3'
H5-fwd	5'-AAATAGGCGTATCACGAGGCC-3'
H5-rev	5'-gcatcagagctcatTTCTGGAATTTTGACTCCTTCC-3'
pQE_upstream-fwd	5'-GTTGAGATCCAGTTCGATGTAACCC-3'
H5_Nde-rev	5'-GGTGATGGTGAGATCCTCTCATatgAGTTAATTTCTCCTCTTTAATG-3'
H5_Nde-fwd	5'-CATTAAAGAGGAGAAATTAACtcatATGAGAGGATCTCACCATCACC-3'

Table 1

Protein expression and purification

H5-SsRG was expressed in *E. coli* ABLE C and purified by His6-tag affinity chromatography: briefly, to remove *E. coli* contaminants, the pool of the eluted fractions from the affinity chromatography was incubated for 10 min at 80 °C, and centrifuged for 30 min at 30,000g. The soluble fraction was dialysed against PBS 1× buffer (phosphate buffer 20 mM, NaCl 150 mM, pH 7.3) and aliquots stored at -20 °C. To assess the purity of the protein samples and determine their concentrations, SDS-PAGE and Bio-Rad protein assay were performed, respectively.

Positive supercoiling assay

Positive supercoiling assay was performed using plasmid pBluescript (Qiagen) as substrate. Briefly, in a total volume of 20 µL, different amounts of protein were incubated with 300 ng of the substrate in 1× reverse gyrase buffer (35 mM Tris-HCl, pH 7.0, 0.1 mM Na₂EDTA, 30 mM MgCl₂, 2.0 mM DTT, 1 mM ATP) at 85°C for 10 min. Samples were analysed by one-dimensional agarose gel electrophoresis. After electrophoresis, gels were stained with ethidium bromide (1 mg/ml), analysed and quantified under UV light with a Chemi-doc apparatus and the QuantityOne software (Bio-Rad, Hercules, CA, USA).

In vitro and in vivo alkyl transferase assay

The activity of the purified fusion protein ogtH⁵-SsRG was measured by a fluorescent-based assay using BG-FL substrate in standard conditions, as previously described (Miggiano et al. 2013; Perugino et al. 2012, 2015). Briefly, in a total volume of 20 µL, 5.0 µM of protein was incubated with 5 µM of the substrate in 1× Fluo Reaction Buffer (50 mM phosphate, 0.1 M NaCl, 1.0 mM DTT, pH 6.5) at 37°C for 1h. Reactions were stopped by denaturing and loading samples on SDS-PAGE, followed by fluorescence imaging analysis on a VersaDoc 4000™ system (Bio-Rad). For the in vivo assay, transformed cells of *S. islandicus* ogt (-) mutant with the pSeSD, the pSeSD-H⁵ and the pSeSD-H⁵-topR1 plasmids were grown at 75 °C in SCV selective medium (0.2% sucrose, 0.2% casamino acids plus a vitamin mixture) as late as stationary phase (O.D.600 nm~ 0.8). Cell pellets from an appropriate volume (typically 1.0 mL) were resuspended in 0.1 mL of SCV medium in the presence of 3.0 µM of BG-FL and incubated at 70 °C for ½ h. After the reaction, cells were first washed twice with 1.0 mL of SCV medium, then spotted on coverslips for the analysis at fluorescence microscopy.

Fluorescence microscopy

A Nikon Eclipse Ti microscope was used to capture the images steps and processed using ImageJ software.

References

1. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. *Science* 1994;263(5148):802–805 (1994).
2. Gronemeyer T, Chidley C, Juillerat A, Heinis C, Johnsson K. Directed evolution of O6-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein Eng Des Sel*. 2006 Jul;19(7):309-16. Epub 2006 Apr 25.
3. Hinner MJ, Johnsson K. How to obtain labeled proteins and what to do with them. *Curr Opin Biotechnol*. 2010 Dec;21(6):766-76. Doi: 10.1016/j.copbio.2010.09.011. Review.
4. Johnsson N, Johnsson K. A fusion of disciplines: chemical approaches to exploit fusion proteins for functional genomics. *Chembiochem*. 2003 Sep 5;4(9):803-10.
5. Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel Johnsson K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* (2003) 21:86–89.
6. Lulchev P, Klostermeier D. Reverse gyrase--recent advances and current mechanistic understanding of positive DNA supercoiling. *Nucleic Acids Res*. 2014 Jul;42(13):8200-13. Doi: 10.1093/nar/gku589. Epub 2014 Jul 10.
7. Miggiano R, Casazza V, Garavaglia S, Ciaramella M, Perugino G, Rizzi M, Rossi F. Biochemical and structural studies of the Mycobacterium tuberculosis O6-methylguanine methyltransferase and mutated variants. *J Bacteriol*. 2013 Jun;195(12):2728-36. doi: 10.1128/JB.02298-12.
8. Napoli,A., Valenti,A., Salerno,V., Nadal,M., Garnier,F., Rossi,M. and Ciaramella,M. Reverse gyrase recruitment to DNA after UV light irradiation in *Sulfolobus solfataricus*. *J. Biol. Chem.*, 279, 33192–33198 (2004).
9. Ogawa T, Sutoh K, Kikuchi A, Kinoshita K Jr. Torsional stress in DNA limits collaboration among reverse gyrase molecules. *FEBS J*. 2016 Apr; 283(8):1372-84. doi: 10.1111/febs.13675.
10. Ogawa T, Yogo K, Furuike S, Sutoh K, Kikuchi A, Kinoshita K Jr. Direct observation of DNA overwinding by reverse gyrase. *Proc Natl Acad Sci U S A*. 2015 Jun 16; 112(24):7495-500. Doi: 10.1073/pnas.1422203112.
11. Perugino G, Miggiano R, Serpe M, Vettone A, Valenti A, Lahiri S, Rossi F, Rossi M, Rizzi M, Ciaramella M. Structure-function relationships governing activity and stability of a DNA

- alkylation damage repair thermostable protein. *Nucleic Acids Res.* 2015 Oct 15;43(18):8801-16. Doi: 10.1093/nar/gkv774.
12. Perugino G, Vettone A, Illiano G, Valenti A, Ferrara MC, Rossi M, Ciaramella M. Activity and regulation of archaeal DNA alkyltransferase: conserved protein involved in repair of DNA alkylation damage. *J Biol Chem.* 2012 Feb 3;287(6):4222-31. Doi: 10.1074/jbc.M111.308320. Erratum in: *J Biol Chem.* 2015 Jan 9;290(2):885.
 13. Tsien RY. The green fluorescent protein. *Annu Rev Biochem* 67:509–544 (1998).
 14. Valenti A, Perugino G, D'Amaro A, Cacace A, Napoli A, Rossi M, Ciaramella M. Dissection of reverse gyrase activities: insight into the evolution of a thermostable molecular machine. *Nucleic Acids Res.* 2008 Aug;36(14):4587-97. Doi: 10.1093/nar/gkn418.
 15. Valenti, A., Napoli, A., Ferrara, M.C., Nadal, M., Rossi, M. and Ciaramella, M. Selective degradation of reverse gyrase and DNA fragmentation induced by alkylating agent in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.*, 34, 2098–2108 (2006).
 16. Valenti, A., Perugino, G., Nohmi, T., Rossi, M. and Ciaramella, M. Inhibition of translesion DNA polymerase by archaeal reverse gyrase. *Nucleic Acids Res.*, 37, 4287–4295 (2009).
 17. Vettone A, Serpe M, Hidalgo A, Berenguer J, del Monaco G, Valenti A, Rossi M, Ciaramella M, Perugino G. A novel thermostable protein-tag: optimization of the *Sulfolobus solfataricus* DNA- alkyl-transferase by protein engineering. *Extremophiles.* 2016 Jan;20(1):1-13. Doi: 10.1007/s00792-015-0791-9.

2.4 A new Daunomycin–Oligoarginine Conjugate: biochemical characterization

Visone V. et al.

Manuscript

Abstract

Cancer is a major health issue that absorbs the attention of a large part of the biomedical research. To fight this disease, new drugs are developed, specifically tailored to target biological pathways or peculiar components of the tumor cells. Particularly interesting in this field is the use of intercalating agents as drugs capable to bind DNA molecules and inhibit enzymes involved in DNA metabolism. Anthracyclines are the most commonly used and effective anticancer drugs. In particular daunomycin (Dau) is used to treat several types of cancers, such as leukemia and neuroblastomas. This agent intercalates in double-stranded DNA and inhibits the cycle of specific DNA topoisomerases, a class of enzymes implicated in the replication processes. Unfortunately, clinical application of anthracyclines is limited by their side effects and the intrinsic or acquired resistance of tumor cells to the drug also reduces the response to the treatment. In order to overcome or reduce these issues, several classes of peptide have been selected as drugs carriers. It has been reported that conjugation of anthracyclines to different types of carrier (e.g. oligo- and polypeptides, proteins, polysaccharides, polymers, and dextran) could improve the selectivity and reduce the side effects by utilizing different cellular uptake mechanisms. Nevertheless the nature of the carrier could influence the chemical and biological properties of conjugate. Miklàn et al., synthesized for the first time a Dau conjugate containing six residues of arginine. However no data about the effect of the carrier on biochemical properties of this compound are available. Here we report a biochemical characterization of Dau-Arg in comparison with the free drug evaluating the fluorescence properties, the DNA-drug interaction and topoisomerase inhibition.

Introduction

Cancer is one of the leading global health problems worldwide with an increasing number of patients every year. The current oncotherapy methods include surgery, chemotherapy, radiotherapy, immunotherapy and the design and synthesis of new drugs for the treatment of cancer is a crucial area of research in medicinal chemistry.

Topoisomerases are ubiquitous nuclear enzymes that catalyze transient cleavage and reconnection of DNA, events which are associated with DNA replication, transcription, recombination and chromatin remodeling. Human cells encode six topoisomerases (TOP1, TOP1mt, TOP2 α , TOP2 β , TOP3 α and TOP3 β), which act on a broad range of DNA and RNA substrates at the nuclear and mitochondrial genomes (Table 1). Their catalytic intermediates, the topoisomerase cleavage complexes, are therapeutic targets of various anticancer drugs (Pommier et al. 2010).

Human topoisomerase	Type	Molecular weight (kDa)	Cleavage	Activities
Top3 α	IA	112	ss	SC- DNA Relaxation
Top3 β	IA	97	ss	SC- DNA Relaxation
Top1	IB	100	ss	SC-/ + DNA Relaxation
Top1mt	IB	70	ss	SC-/ + DNA Relaxation
Top2 α	IIA	170(x2)	ds	SC-/ + DNA Relaxation DNA Decatenation
Top2 β	IIA	180(x2)	ds	SC-/ + DNA Relaxation DNA Decatenation

Table 1: classification of human DNA topoisomerases. ss: single strand DNA; ds: double strand DNA; sc: supercoiled DNA; (x2):dimeric.

Anthracyclines (e.g. doxorubicin, daunomycin, idarubicin etc) are a group of antibiotics, discovered over 50 years ago, which exhibit cytostatic activity and are widely used in the treatment of various types of cancer, such as osteosarcomas and acute myeloid leukemia (Minotti et al. 2004). Anthracyclines structure consists of a tetracycline ring (4 rings connected to each other) with a sugar moiety (called daunosamine) attached by a glycosidic linkage. Despite the broad and relatively established usage of anthracyclines in antitumor therapy, their mechanism of action is still not fully understood.

Daunomycin (Dau) is one of the anthracycline-type antitumor agents able to inhibit the cells division by intercalating to DNA or/and by inhibiting the topoisomerase II (Fig 1).

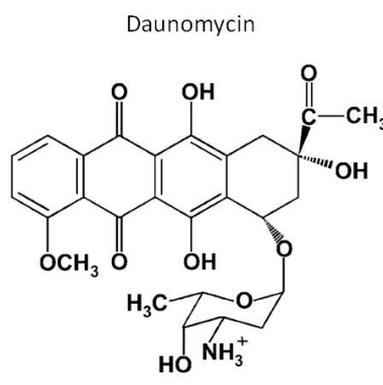


Fig.1: chemical structure of Daunomycin.

In cancer cells the inhibition of topoisomerase II causes the formation of the Dau-topoisomerase II complex, resulting in DNA fragmentation, inhibition of cellular proliferation and finally cell death. Leading to the apoptosis is probably the dominant mechanism of antitumor action of the anthracyclines (Zunino et al. 1990; Gewirtz et al. 1999; Ciesielska et al 2005).

The anthracyclines are largely used in chemotherapy; however clinical application of anthracyclines is limited by their side effects. Furthermore the intrinsic or acquired resistance of tumor cells to anthracyclines also reduces the response to the treatment (Von Hoff et al. 1977).

In order to reduce the side-effects, to circumvent multidrug resistance, to increase selectivity or target the molecule to specific part of the body several classes of peptide have been selected as drugs carriers. They are water-soluble and undergo facilitate uptake into cells and tissues. Among these molecular transporters, short oligomers of arginine containing five or more Arg residues are considered as cell-penetrating peptides capable to translocate covalently attached cargo into the cytosol (Miklàn et al. 2009).

Nevertheless, several lines of evidence suggested that nature of the carrier could influence the chemical and biological properties of conjugate, therefore is essential to evaluate the specific contribution of the molecular transporter for each compound.

Previously, Miklàn et al. (2009) described for the first time the synthesis of Dau-conjugate in which an oligoarginine with 6 residues is attached to the drug by an oxime bond (Dau-Arg₆, Fig. 2).

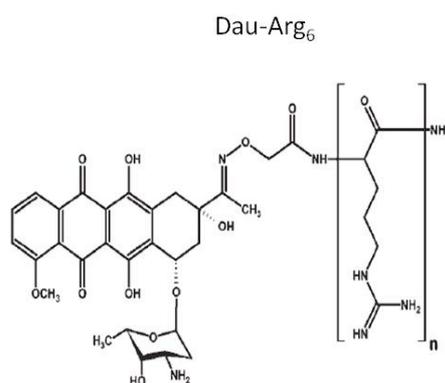


Fig.2: chemical structure of Dau-Arg₆ from miklàn et al. (2009).

They demonstrated that Dau-Arg₆ conjugate is stable in a broad range of conditions (in distilled water as well as in 0.1M Na-citrate/citric acid buffer at pH=5 and pH=2) and also when kept at 4°C in solid form; the typical fluorescent spectrum related to drug moiety is not influenced by the presence of the hexaarginine; it retains the cytostatic effect on viability of HL-60 and HepG2 cell; finally it is taken up more efficiently by HL-60 than HepG2 cells as already known for the free drug.

Taken together, these results represent a good assumption to the rational design of improved antitumor agents and spurred us to further investigate about the biochemical properties of this new conjugate.

Results and Discussion

Effect of Dau-Arg on DNA relaxation activity of human topoisomerases

DNA topoisomerases relax torsional stress formed during biological processes such as DNA replication and transcription. Specifically the enzymes cut the DNA in one or both strands, depending upon whether DNA topoisomerase II or I is involved, pass the other duplex through the resulting gap and reseal the break (Fig. 3).

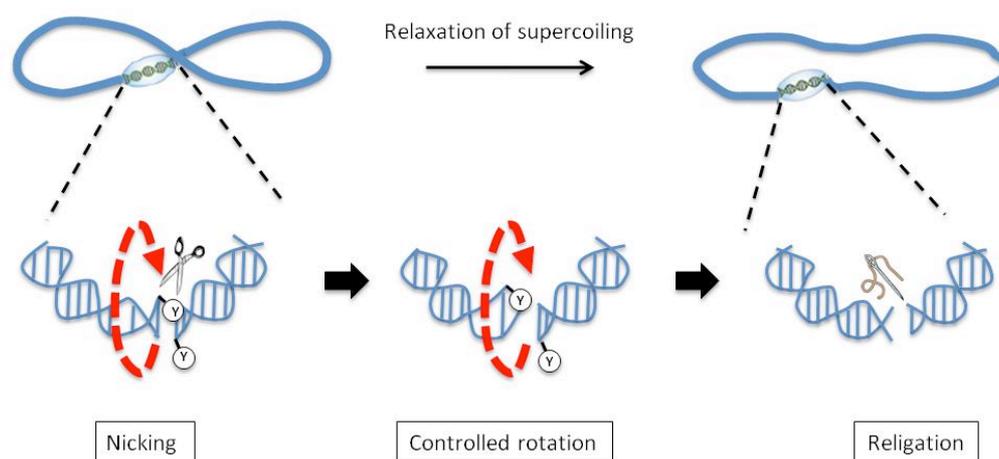


Fig.3: TopoII-mediated DNA relaxation by controlled rotation. This reaction requires ATP as energy cofactor and divalent metal cations. Adapted from Pommier et al. (2010).

By forming a drug-enzyme-DNA complex, the topoisomerase inhibitors prevent the subsequent DNA-resealing step normally catalyzed by topoisomerase. The result is an inhibition of relaxation activity (Fig. 4).

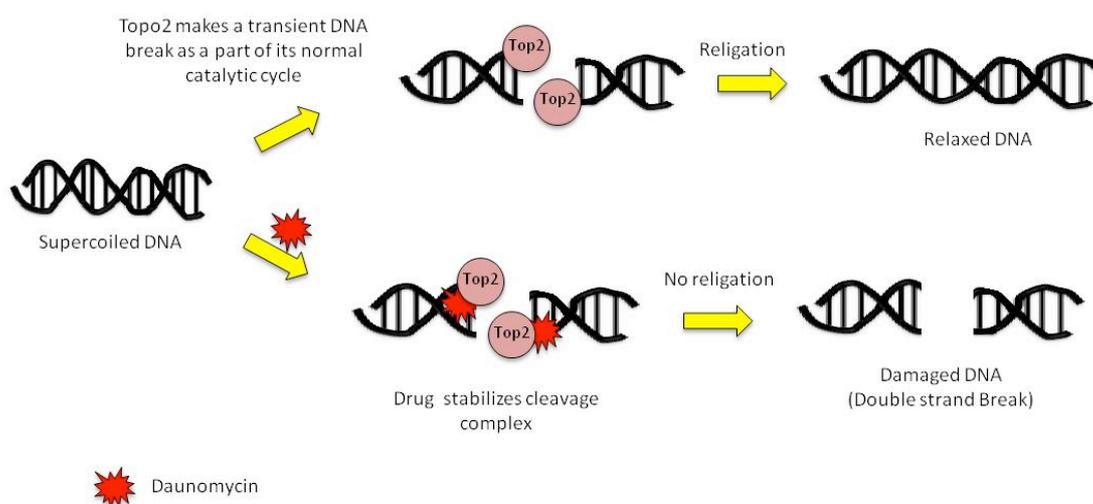


Fig.4: trapping of the cleavage complex by Daunomycin.

In order to evaluate the contribute of Arginine carrier on Dau-coniugate we tested the inhibition of topoisomerase activity by coniugate in comparison with the free drug. For this purpose we performed a

relaxation assay using supercoiled plasmid as substrate of human topoisomerase II α , the isoform strongly associated with cell growth and division.

The principal approach to evaluate the topoisomerases activity and the drug inhibition is based on electrophoretic properties of supercoiled DNA that has a different mobility than completely relaxed DNA (Fig. 5A).

In Fig. 5B the gel shows the different electrophoretic mobility properties of plasmid before (lane 1) and after incubation with increasing concentrations of purified human topoisomerase II (lanes 2-5). The relaxation of the plasmid resulting in a range of products that migrate slower than the supercoiled one.

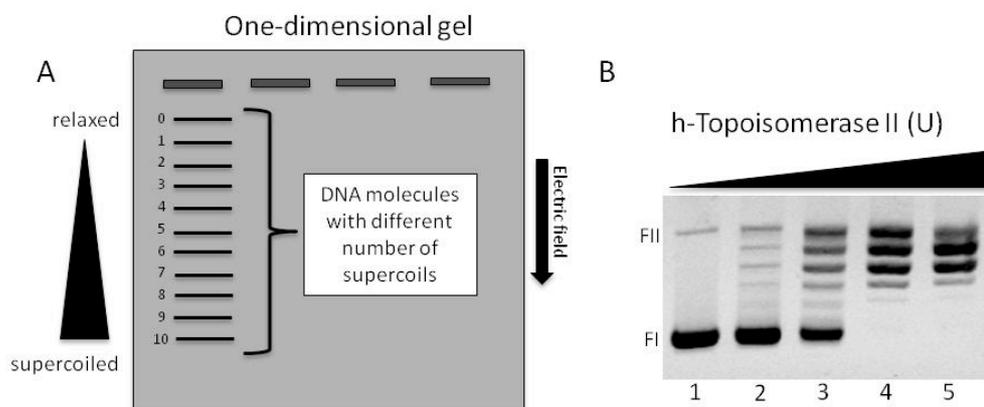


Fig.5: panel A shows a schematic representation of gel migration of topoisomers from supercoiled to relaxed forms. Panel B shows relaxation activity of human topoisomerase II α . Plasmid was incubated without (lane 1) or with (lane 2-5) increasing concentrations of hTopoII α (0.1, 0.2, 0.5 and 1 Units). The relative mobilities of negatively supercoiled plasmid DNA (form I, FI) and nicked circular plasmid DNA (form II, FII) are shown.

In the gel in Fig. 6A Daunomycin and its Arg-derivative were added to reactions containing a fixed topoisomerase II concentration at indicated amounts. Under the conditions used, inhibition starts at 1 micromolar of daunomycin and is complete at 5 micromolar. Interestingly the Dau-Arg₆ conjugate inhibits topoisomerase II activity in a concentration-dependent manner similarly to Dau.

In order to quantify the drugs inhibition, a densitometric analysis of the topoisomers band was performed and therefore the residual topoisomerase II activity was determined (Fig. 6B).

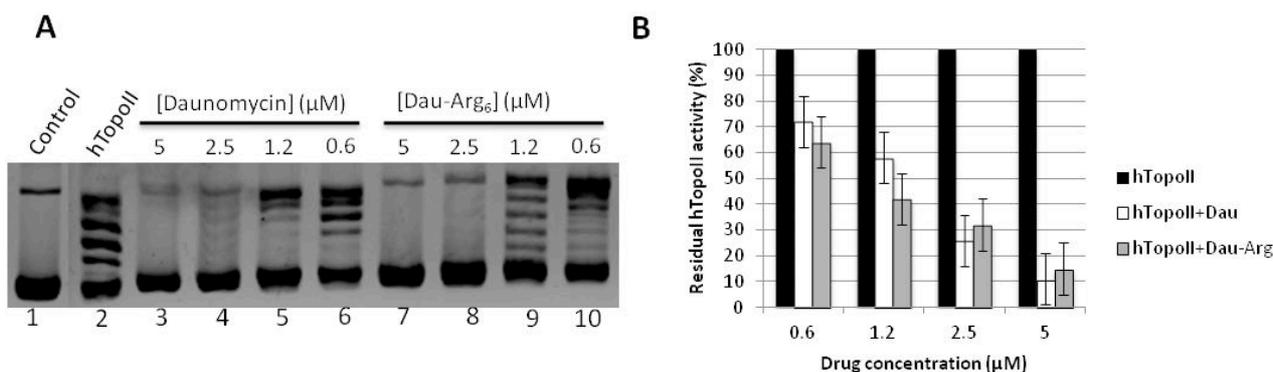


Fig.6: Dau and Dau-Arg inhibit DNA relaxation catalyzed by human topoisomerase II. Panel A shows the effects of both drugs on DNA relaxation catalyzed by hTopoII. Plasmid was incubated without (lane 1) or with (lane 2-10) hTopoII (0.4 U). Different amounts of drugs were added as indicated (lane 3-10). Panel B: quantification of hTopoII activity as a function of drug concentration. Error bars represent standard deviations of 3 independent assays.

Like the free drug, at 1.2 μM of the Dau-Arg the residual topoisomerase activity was about 50%, whereas at 5 μM only 10% was observed, confirming that the hexaarginine peptide does not influence the topoisomerase-inhibitory effect of the Daunomycin.

We also tested the Dau-Arg₆ effect on Topoisomerase I another target of Daunomycin. In contrast to TopoII, the mechanism of inhibition of TopoI does not involve the trapping of DNA covalent complex, however inhibition of DNA relaxation also occurred. Failure of anthracyclines to stabilize the covalent complex of the topoisomerase I reaction suggests that the inhibition is not specific, arising solely from DNA binding rather than from interaction between drug and covalent enzyme-DNA species.

As show in Fig. 7(A and B) the Dau-Arg₆ has a similar inhibitory effect of Daunomycin confirming that the oligo-arginine does not influence the inhibition of Topoisomerase I activity by DNA binding.

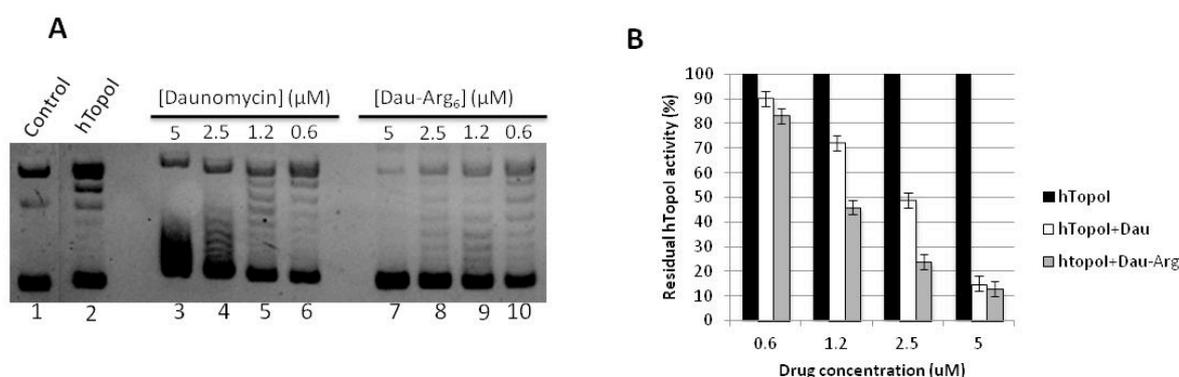


Fig.7: Dau and Dau-Arg inhibit DNA relaxation catalyzed by human topoisomerase I. Panel A shows the effects of both drugs on DNA relaxation catalyzed by hTopoI. Plasmid was incubated without (lane 1) or with (lane 2-10) hTopoI (100 nM). Different amounts of drugs were added as indicated (lane 3-10). Panel B: quantification of hTopoI activity as a function of drug concentration. Error bars represent standard deviations of 3 independent assays.

Spectroscopic analysis of DNA-drug interaction

Despite the inhibition of topoisomerases by anthracyclines is believed to contribute to their efficacy against cancer cells, the mechanism by which these drugs block the function of enzymes is not completely understood. The biological activity of Daunomycin is likely associated with its DNA binding properties and there is some evidence that antitumor activity is essentially due to the intercalation in the base pairs of DNA (Calendi et al. 1965) (Fig. 8).

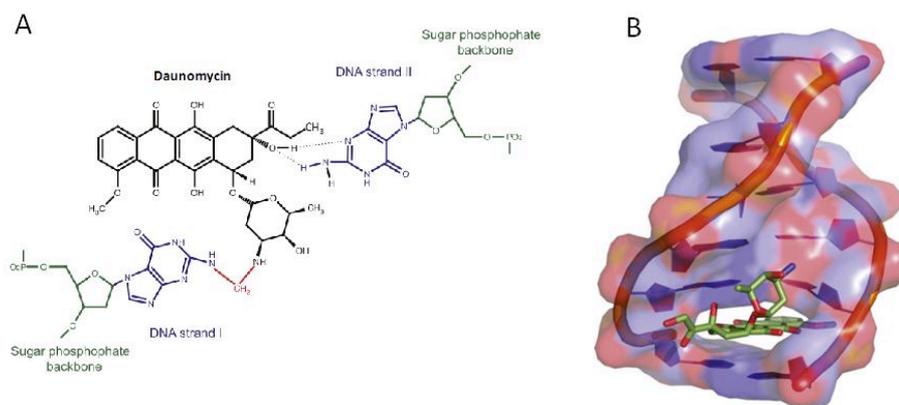


Fig.8: (A) chemical representation of interaction between Dau and DNA. (B) Three-dimensional model of Dau-DNA interaction.

In order to evaluate the contribute of Arginine on DNA-drug interaction we performed an in-deep analysis of DNA binding mediated by Dau and Dau-Arg₆ using different experimental approaches.

- **Fluorescence measurements**

Intercalation of anthracyclines into double helix of DNA can be easily monitored by fluorescence spectroscopy. Dau is a fluorescent compound that shows an emission spectrum in the 500–700 nm range when excited at 480 nm (Fig. 9).

We also checked the fluorescence spectrum of Dau-Arg₆ showing the same trend of the free drug (Fig. 9).

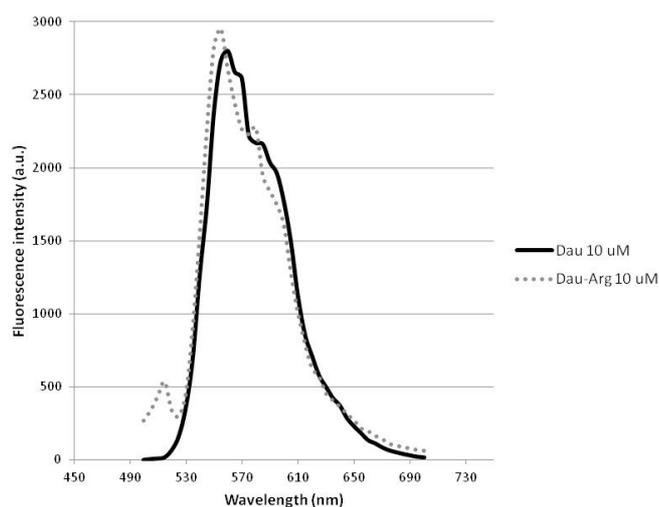


Fig.9: emission spectra of Dau and Dau-Arg conjugate using $\lambda= 480$ nm excitation and recorded at 10 μ M concentration in Tris buffer (pH 7.5) containing 20 mM Tris-HCl.

As showed in Fig 10A, increasing concentrations of DNA plasmid to Daunomycin, reduce the fluorescence emission of the drug suggesting that the DNA binding quench the fluorescence intensity of Dau. In order to analyze the interaction between DNA and Dau-Arg₆, we evaluate the fluorescence quenching of the conjugate in the presence of DNA. Interestingly when increased amounts of DNA are added, a decrease of fluorescence emission peak was observed, indicating a strong association of this drug with DNA as already showed for Dau (Fig. 10B).

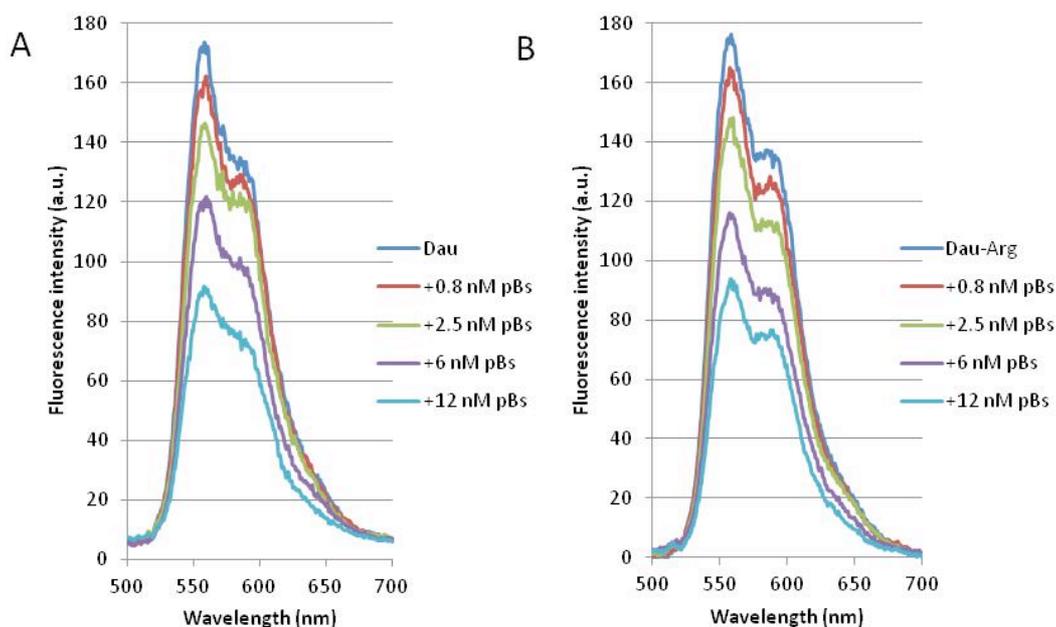


Fig.10: emission spectra of Dau (A) and Dau-Arg conjugate (B) in Tris-HCl buffer (pH 7.5) in absence and presence of increasing concentrations of pBluescript DNA (pBs).

- **Circular dichroism analysis**

Having obtained preliminary information on DNA-drug interaction by fluorescence spectroscopy the DNA binding was further studied by circular dichroism (CD) analysis. Due to its accuracy and sensitivity, circular dichroism measurement is commonly used to monitor the conformational changes of DNA structure. Furthermore this technique can be used to investigate the changes in DNA morphology and to determine the mode of drug-DNA interaction. The changes in CD signals of B-DNA, as observed after drugs binding can be assigned to the corresponding changes in DNA morphology. In order to evaluate the interaction between the drugs and DNA we performed circular dichroism analysis using calf thymus DNA (ctDNA) as substrate.

In Fig. 11 is reported the CD spectrum of calf thymus DNA showing the typical features of B-DNA: a positive band at 275 nm due to the base stacking and a negative band at 245 nm specific of DNA helicity. It is known that interaction between DNA and molecules induce variation in these bands. Specifically, whereas binding to DNA grooves result in little or no changes of the base stacking and helicity bands, intercalation of molecules into the double helix induces a noteworthy change of intensity of the bands (Kypr et al. 2009; Sprecher et al. 1979).

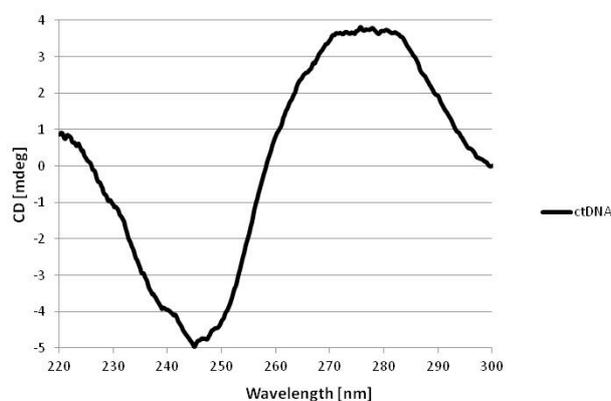


Fig.11: CD spectrum of 0.2 mg/mL ctDNA at 20°C in 10 mM Tris-HCl (pH 7.5).

Is reported that in the range of 320-200 nm the spectrum changes occurring upon addition of the anthracyclines to DNA are related to structural modification of the nucleic acid (Cera et al. 1991).

In our conditions the CD spectra of Daunomycin and Dau-Arg₆ show a changes of negative band at 245 nm (shifting to zero level) suggesting that both drugs modify DNA structure by helix intercalation (Fig. 12). Notably, a difference between the CD spectra of DNA in presence of the two drugs is observed at 275 nm, where a decrease of the band was observed with Dau- Arg₆, but not with Dau, indicating a perturbation of the bases stacking interactions. Is reported that the DNA spectrum in presence of Arginine shows a decrease in 275 nm peak intensity with arginine concentration suggesting that hydrophobic and electrostatic forces are involved in DNA-Arg binding (Arakawaa et al. 2010). Our

results suggest that electrostatic binding of ions could be also involved in Dau-Arg-DNA interaction.

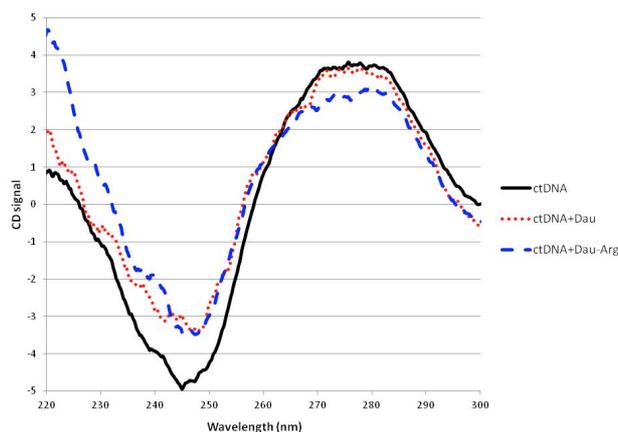


Fig.12: CD spectra of ctDNA (0.2 mg/mL) in 10 mM Tris-HCl (pH 7.5) in presence of 4 μ M of Dau and Dau-Arg.

The Induced Circular Dichroism (ICD) experiment is also helpful to evaluate the binding modes of small molecules to DNA. Many DNA binding ligands are achiral and so optical inactive. However, through interaction with DNA, these ligands can acquire an induced CD signal. The observation of an ICD signal within the absorption bands of the achiral ligand is immediately indicative of a ligand–DNA interaction (Garbett et al. 2007). It is well known that a positive ICD peak is due to groove binding and a negative ICD peak is due to intercalative binding (Bayley et al. 1969; Tidwell et al. 2003).

In our hands upon DNA addition the ICD spectra of both drugs shows a minimum at about 300 nm as also known for other anthracyclines (Airoldi et al 2014) (Fig. 13) suggesting an intercalative binding mode of Dau and its derivative.

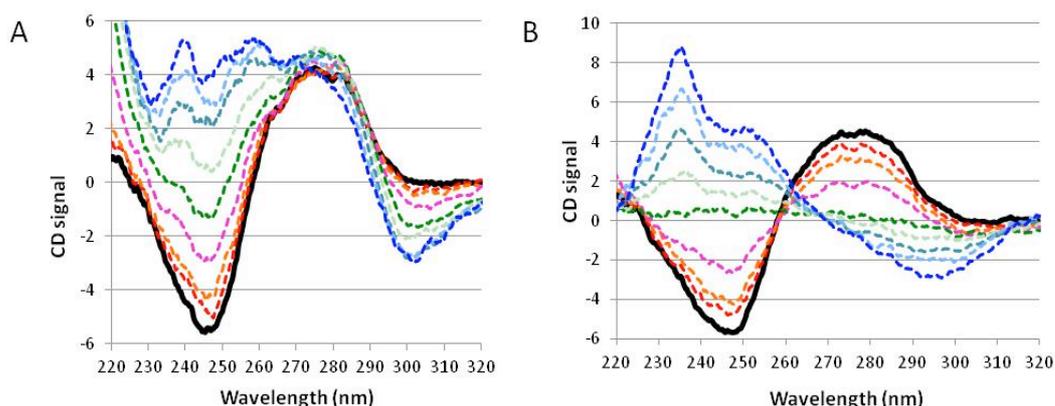


Fig.13: titration of ctDNA with Dau and Dau-Arg. (A) CD spectra of titration of ctDNA with Dau. (B) spectra of titration of ctDNA with Dau-Arg. Black curve is the spectrum of ctDNA alone (0.2 mg/mL); dashed lines are the increasing concentrations of Dau and Dau-Arg (--- 5 μ M --- 10 μ M --- 20 μ M --- 40 μ M --- 60 μ M --- 80 μ M --- 100 μ M --- 120 μ M).

Our results indicate that, like Dau, the oligo Arg derivative interacts with DNA mainly intercalating into the double helix.

• Melting studies

Thermal denaturation of DNA was considered a direct method to determine the stabilization/destabilization effect of ligands on DNA double helix (Wilson et al. 1981; Patel et al. 2010). In order to perform an in deep analysis of DNA-Dau Arg₆ interaction, melting experiments were carried out monitoring the melting temperature of DNA with or without compounds by circular dichroism. The Fig. 14 shows the melting profiles of ctDNA in absence and in presence of the drugs. Without the drugs, a temperature dependent base pairs opening was observed and the calculated temperature of melting (T_m) was about 81°C. The presence of Dau or Dau-Arg₆ induces a significant increase of melting temperature by about 9°C confirming the intercalative binding mode of both Dau and its oligo-Arg derivative.

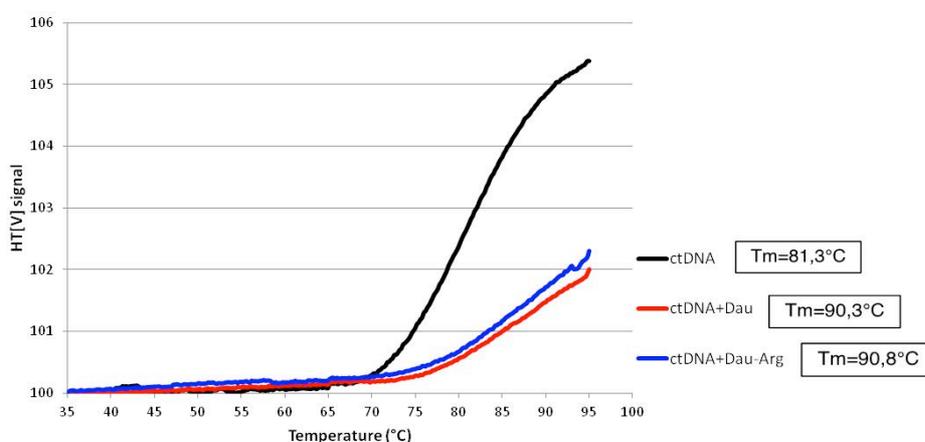


Fig.14: melting curves of ctDNA (0.2 mg/mL) at 260 nm in absence and presence of 4 μ M Dau and Dau-Arg in Tris-HCl buffer (pH 7.5). The melting temperatures (T_m) were determined as described in Materials&Methods.

Taken together the studies of DNA-drugs interaction demonstrate that the Dau-Arg₆ conjugate is able to interact and stabilize DNA helix as well as free drug suggesting that the presence of the hexaarginine-peptide does not influence the DNA binding properties of Daunomycin.

Conclusion

Daunomycin is widely used in cancer chemotherapy. Despite its usage is well established, a lot of side effects limited its involvement in antitumor therapy. Further development of its efficacy could cover the synthesis of analogues and conjugates for tumor cell targeting. In a previous work, a new Dau-Arg conjugate was prepared and comparative investigation was performed. An oligoarginine peptide was selected as carrier for conjugation. It has been reported that oligoarginines containing five or more arginine residues are considered as cell-penetrating peptides capable to translocate covalently attached cargo into the cytosol (Derossi et al. 1994; Fawell et al. 1994). The analysis of cell-penetrating properties is in the focus of intense research but the mechanism by which these compounds enter the cells is still not completely clear (Miklàn et al. 2009).

In the present study, we examined the biological properties of Daunomycin-Arg₆ conjugate on double helical structure of DNA using biochemical and spectroscopic techniques. Relaxation assays show that Daunomycin-Arg₆ conjugate inhibits the human Topoisomerase II activity in a concentration-dependent manner similarly to Dau. The Daunomycin-Arg₆ conjugate exhibits a high binding affinity for ctDNA. Different methods were used to investigate the interaction mechanism. Fluorescent quenching assay confirms that binding of Dau-Arg₆ with DNA occurs mainly by intercalation. Indeed, addition of increasing amounts of ctDNA to drug shows a decrease in Dau-Arg₆ fluorescent emission. CD results showed deep conformational changes in the ctDNA double helix upon binding with the drug. In our experiments the CD spectra obtained with both Daunomycin and Dau-Arg₆ showed changes of negative band at 245 nm, suggesting that both drugs intercalate in the double helix causing a structural distortion of the canonical B-conformation (as already known for Dau, Quadrifoglio et al 1974; Suh et al 2002; Ray et al 2003).

It is known that intercalation of small molecules into the double helix increases the helix melting temperature and stabilizes the natural structure of DNA, while non-intercalative binding causes no significant variation of T_m . The melting curves of ctDNA in presence of both Dau and Dau-Arg₆ show an increase of the melting temperature by about 9 °C, confirming the intercalative binding mode of both Dau and its oligoarginine derivative. In conclusion, our results indicate that, like Daunomycin, the oligoarginine derivative interacts with DNA mainly intercalating into the double helix. Our observations also suggest that the Arg-moiety is involved in DNA binding properties of the conjugate. This study is expected to provide greater insight into the use of this new class of Dau-conjugates as anticancers.

Materials and Methods

Materials

Calf thymus DNA (catalog number D1501) was purchased from Sigma and dissolved in 10mM Tris-HCl, pH 7.5 at 1 mg/ml. The reconstituted solution was stored at -20 °C until experiments. The DNA sample at 1mg/ml was diluted to either 0.2 mg/ml for CD measurements in 10mM Tris-HCl, pH 7.5. Daunomycin and its Arg-derivative were prepared as previously reported by Miklàn et al. (2009) and kept at 4°C. Human topoisomerase II α and 10X Topoisomerase II Reaction Buffer (product number 78303) were purchased by Affymetrix and stored at -20 °C. Human topoisomerase I was a kind gift of Prof. Piero Benedetti. pBluescript negatively supercoiled plasmid was directly extracted from an overnight *E. coli* cell culture by Invitrogen HiPure Plasmid Midi prep Kit.

Topoisomerase-Mediated Relaxation of Plasmid DNA

Reaction mixtures contained 0.4 units of topoisomerase II α , 7.5 nM negatively supercoiled pBluescript DNA, and 1 mM ATP in 20 μ L of 10 mM Tris-HCl, pH 7.9, 50 mM NaCl, 50 mM KCl, 0.1 mM EDTA, 5 mM MgCl₂, and 15 μ g/ml BSA. Assays were started by the addition of enzyme, and DNA relaxation mixtures were incubated for 30 min at 30 °C. For human topoisomerase I reaction mixtures contained

100 nM of the enzyme, 7.5 nM negatively supercoiled pBluescript DNA in 20 μ L of 10 mM Tris-HCl, pH 7.0, 100 mM NaCl, 0.1 mM NaEDTA, 10 mM MgCl₂ and 50 μ g/ml BSA and were incubated for 1 h at 37°C. DNA relaxation reactions were carried out in the presence of 0–5 μ M Dau and Dau-Arg. In order to determine the ability of Dau and Dau-Arg to inhibit DNA relaxation activity, reactions were incubated for 10 min in ice prior to the addition of topoisomerases. Control experiments were performed in the absence of the compounds to compare enzyme activity. DNA relaxation was stopped by the addition of 2 μ L of 10% SDS. Samples were subjected to gel electrophoresis in 1.2% agarose gels. The agarose gel was then stained in ethidium bromide for 10 min. DNA bands were visualized using a BioRad Versadoc imaging system and quantified by Quantity One® software. DNA relaxation was monitored by the conversion of supercoiled plasmid DNA to relaxed topoisomers.

Fluorescence spectroscopic studies

All fluorescence measurements were carried out with a JASCO spectrofluorometer (FP8200). Spectra of Dau and its derivative were recorded at 10 μ M concentration in Tris buffer (pH 7.5) containing 20 mM Tris-HCl. For quenching experiment the concentration of Dau-Arg (10 μ M) was kept constant while the DNA concentration was varying from 5 to 40 nM. The excitation wavelength was $\lambda = 480$ nm, and the emission was detected between 500 and 700 nm.

CD measurements

CD measurements were carried out on a Jasco J-810 spectropolarimeter at controlled temperature. A Peltier cell holder and a Julabo F-25 temperature programmer were used to regulate the temperature of the cell. A 0.1 cm cell was used for 0.2 mg/mL sample. Following parameters were used: 20 nm/min scan speed, 3 accumulations, 4 s time constant, 1 nm band width and 0.2 nm data pitch for spectral measurements and 1°C/min temperature increment for melting measurements. The CD spectra of ctDNA in the absence and presence of 4 μ M Dau and its conjugate Dau-Arg were measured in pH 7.5 Tris-HCl 10 mM buffer under a nitrogen atmosphere. All observed CD spectra were an average of three scans recorded at a speed of 20 nm/min and corrected for buffer signal. ICD measurements were recorded by keeping the concentration of ctDNA constant (0.2 mg/mL) and increasing the drug concentration from 0 to 120 μ M.

Melting studies

Thermal melting of DNA was followed at 260 nm by HT[V] signal. HT[V] signal show a gradual change with temperature in 10mM Tris-HCl buffer, pH 7.5: upon melting the HT[V] signal increases due to loss of base stacking. HT[V] may be more sensitive than CD in seeing melting of DNA. The temperature of melting, T_m , was determined from each transition data by fitting sigmoidal curves using Boltzmann equations and obtaining good R squared values (>0.95).

References

1. Airoidi M, Barone G, Gennaro G, Giuliani A M and Giustini M. Interaction of Doxorubicin

- with Polynucleotides. A Spectroscopic Study. *Biochemistry* 2014, 53, 2197–2207.
2. Arakawaa T, Hiranob A, Shirakib K, Kitac Y, Koyamad A H. Stabilizing and destabilizing effects of arginine on deoxyribonucleic acid. *International Journal of Biological Macromolecules* 46 (2010) 217–222.
 3. Bayley P M, Nielsen E B and Schellman J A. The rotatory properties of molecules containing two peptide groups: theory. *J. Phys. Chem.* 73, 228–243 (1969).
 4. Calendi E, Di Marco A, Reggiani M, Scarpinato B and Valentini L. On physico-chemical interactions between daunomycin and nuclerc acids. *Biochim. Biophys. Acta*, 103 (1965) 25–49.
 5. Cera C, Palù G, Magno SM, Palumbo M. Interaction between second generation anthracyclines and DNA in the nucleosomal structure. *Nucleic Acids Res.* 1991 May 11;19(9):2309-14.
 6. Chaires J B, Dattagupta N and Crothers D M. Studies on Interaction of Anthracycline Antibiotics and Deoxyribonucleic Acid: Equilibrium Binding Studies on Interaction of Daunomycin with Deoxyribonucleic Acid? *Biochemistry* 1982, 21, 3933-3940.
 7. Ciesielska E, Studzian K, Wasowska M, Oszczapowicz I, Szmigiero L. Cytotoxicity, cellular uptake and DNA damage by daunorubicin and its new analogues with modified daunosamine moiety. *Cell Biol. Toxicol.* 21 (2005) 139–147, <http://dx.doi.org/10.1007/s10565-005-0142-1>.
 8. Derossi D, Joliot A H, Chassaing G; Prochiantz A J. The third helix of the Antennapedia homeodomain translocates through biological membranes. *Biol Chem* 1994, 269, 10444–10450.
 9. Fawell, S, Seery J, Daikh Y, Moore C, Chen L L, Pepinsky B, Barsoum J. Tat-mediated delivery of heterologous proteins into cells. *Proc Natl Acad Sci USA* 1994, 91, 664–668.
 10. Garbett N C, Ragazzon P A and Chaires J B. Circular dichroism to determine binding mode and affinity of ligand–DNA interactions. *Nature Protocols* VOL.2 NO.12 2007.
 11. Gewirtz D.A. A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. *Biochem. Pharmacol.* 57 (1999) 727–741.
 12. Kypr J, Kejnovská I, Renčiuk D, Vorlíčková M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* 2009 Apr; 37(6): 1713–1725.
 13. Mikla'n Z, Orba'n E, Csí'k G, Schlosser G, Magyar A, Hudecz F. New Daunomycin–Oligoarginine Conjugates: Synthesis, Characterization, and Effect on Human Leukemia and Human Hepatoma Cells. *PeptideScience* Volume 92 / Number 6 (2009) DOI 10.1002/bip.21264.
 14. Minotti G, Menna P, Salvatorelli E, Cairo G, Gianni L. Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. *Pharmacol Rev.*

2004 Jun;56(2):185-229.

15. Patel M N, Dosi P A, Bhatt B S. Antibacterial, DNA interaction and superoxide dismutase activity of drug based copper (II) coordination compounds. *Polyhedron* 29 (2010) 3238–3245.
16. Pommier Y, Leo E, Zhang H and Marchand C. DNA Topoisomerases and Their Poisoning by Anticancer and Antibacterial Drugs. *Chemistry & Biology* 17, May 28, 2010.
17. Quadrifoglio F, Crescenzi V. On the binding of actinomycin and of daunomycin to DNA: a calorimetric and spectroscopic investigation. *Biophys Chem.* 1974 Jun;2(1):64-9.
18. Ray A, Kumar GS, Maiti M. Molecular aspects on the interaction of aristololactam-beta-D-glucoside with H(L)-form deoxyribonucleic acid structures. *J Biomol Struct Dyn.* 2003 Aug;21(1):141-51.
19. Sprecher CA, Baase WA, Johnson WC Jr. Conformation and circular dichroism of DNA. *Biopolymers.* 1979 Apr;18(4):1009-19. DOI: 10.1002/bip.1979.360180418.
20. Suh D, Oh YK, Ahn B, Hur MW, Kim HJ, Lee MH, Joo HS, Auh C. Comparative binding of antitumor drugs to DNA containing the telomere repeat sequence. *Exp Mol Med.* 2002 Nov 30;34(5):326-31.
21. Szafraniec E, Majzner K, Farhane Z, Byrne H.J., Lukawska M, Oszczapowicz I, Chlopicki S, Baranska M. Spectroscopic studies of anthracyclines: Structural characterization and in vitro tracking. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 169 (2016) 152–160.
22. Tidwell R R and Boykin D W. Dicationic DNA minor groove binders as antimicrobial agents. *DNA and RNA Binders: From Small Molecules to Drugs Vol. 2* (eds. Demeunynck, M., Bailly, C. & Wilson, W.D.) 416–460 (Wiley-VCH, Weinheim, Germany, 2003).
23. Von Hoff DD, Rozencweig M, Layard M, Slavik M, Muggia FM. Daunomycin-induced cardiotoxicity in children and adults. A review of 110 cases. *Am J Med.*1977 Feb;62(2):200-8.
24. Zunino F, Capranico G. DNA topoisomerase II as the primary target of anti-tumor anthracyclines. *Anticancer Drug Des.* 5 (1990) 307–317.

2.5 Chromatin Structure and Dynamics in Hot Environments: Architectural Proteins and DNA Topoisomerases of Thermophilic Archaea

Valeria Visone, Antonella Vettone, Mario Serpe, Anna Valenti, Giuseppe Perugino, Mosè Rossi and Maria Ciaramella *

Institute of Biosciences and Bioresources, National Research Council of Italy, Naples 80131, Italy;
E-Mails: valeria.visone@ibbr.cnr.it (V.V.); antonella.vettone@ibbr.cnr.it (A.Ve.);
mario.serpe@ibbr.cnr.it (M.S.); anna.valenti@ibbr.cnr.it (A.Va.);
giuseppe.perugino@ibbr.cnr.it (G.P.); mose.rossi@ibbr.cnr.it (M.R.)

* Author to whom correspondence should be addressed; E-Mail: maria.ciaramella@ibbr.cnr.it;
Tel.: +39-81-6123-246; Fax: +39-81-6123-646.

Received: 22 July 2014; in revised form: 19 August 2014 / Accepted: 9 September 2014 /

Published: 25 September 2014

Abstract: In all organisms of the three living domains (Bacteria, Archaea, Eucarya) chromosome-associated proteins play a key role in genome functional organization. They not only compact and shape the genome structure, but also regulate its dynamics, which is essential to allow complex genome functions. Elucidation of chromatin composition and regulation is a critical issue in biology, because of the intimate connection of chromatin with all the essential information processes (transcription, replication, recombination, and repair). Chromatin proteins include architectural proteins and DNA topoisomerases, which regulate genome structure and remodelling at two hierarchical levels. This review is focussed on architectural proteins and topoisomerases from hyperthermophilic Archaea. In these organisms, which live at high environmental temperature (>80 °C <113 °C), chromatin proteins and modulation of the DNA secondary structure are concerned with the problem of DNA stabilization against heat denaturation while maintaining its metabolic activity.

Keywords: DNA topology; DNA structure; thermophilic organisms

1. Introduction

Chromatin associated proteins have the essential function of compacting, shaping and modelling the genome structure. Thanks to their combined action, the genome is organized into higher order, highly

regulated and dynamic structures, which reduce its enormous length to fit into the nuclear or nucleoid compartment, and make complex genome functions possible. Indeed, chromatin structure influences all information processes (transcription, replication, recombination and repair) and chromatin remodelling plays important regulatory roles in all these processes.

Chromatin proteins include architectural proteins and DNA topoisomerases. The first are small, basic DNA-interacting proteins generally not conserved at the primary sequence level, whose binding mode and structural effects on the genome are similar. They can induce DNA bending, looping, bridging or wrapping [1]. DNA topoisomerases are essential and evolutionary highly conserved enzymes inducing covalent modifications of DNA secondary structure and are responsible for the maintenance of proper DNA topology during the entire life of the cell [2–4]. Both classes of proteins contribute to maintenance and modulation of genome structure at two hierarchical levels, which affect each other; for example, the interaction between an architectural protein and its binding site on DNA can be regulated by DNA secondary structure and, conversely, DNA topoisomerases activity may be regulated by DNA binding proteins.

Chromatin structure studies, in particular in eukaryotes, have recently received great support thanks to the resolution of a number of 3D structures of these proteins and their complexes with DNA, as well as to the development of highly sophisticated methods employing atomic force microscopy, optical and magnetic tweezers, fluorescence imaging and chromatin sequencing. These techniques allow a wide range of analyses, from single-molecule up to genome-wide level, addressing the mechanisms and details of chromatin structure and function *in vitro* and *in vivo* (see for instance: [5–9]).

The Archaea comprise procaryotic microorganisms forming an evolutionary and functional domain distinct from Bacteria and Eukaryotes. Many archaeal species are characterized by peculiar and extreme habitats (hot springs, deep hydrothermal vents, saline and alkaline water, acid mines, antarctic ice, and so on). Although there is little knowledge of chromatin structure in Archaea, it is clear that their genomes are organized into a compact nucleoid. We here focus on architectural proteins and topoisomerases from hyperthermophilic archaea. In these organisms, which live at high environmental temperature (>80 °C <113 °C), chromatin proteins have the additional task of protecting DNA from denaturation while maintaining the flexibility needed to allow information processes. Most proteins and enzymes from these organisms show intrinsic high stability to heat and high thermophilicity, and generally their optimal temperature ranges are consistent with the growth temperature of their source. Some of them show the same activities as their mesophilic counterparts, but with higher thermal stability and higher ranges of temperature optima; other show very peculiar activities not found in proteins from other organisms. While we have tried to give a wide overview of the chromatin field in hyperthermophilic archaea and summarize main recent results, we are aware we could not cover many important aspects of this topic; we apologize to colleagues whose work was not cited and direct readers to a number of excellent more specialized reviews [10–16].

2. Architectural Proteins of Hyperthermophilic Archaea

Archaea include at least two well-studied kingdoms, the Euryarchaea and Crenarchaea, as well as three other less well-studied groupings, the Nanoarchaea, Korarchaea, and Thaumarchaea. Most studies on chromatin proteins have been performed on members of Euryarchaea and Crenarchaea, which show considerable diversity in chromatin-associated proteins: whereas Euryarchaea encode proteins similar to eukaryotic histones, most Crenarchaea typically do not, and instead contain a set of different architectural DNA-binding proteins [11–13]; some of these are shared by the two groups and have homologs in organisms outside the archaeal domain, but others are unique to one kingdom or even one genus (Table 1). This diversity is quite puzzling, also considering that information processing pathways show striking structural and functional conservation from archaea and eukaryotes.

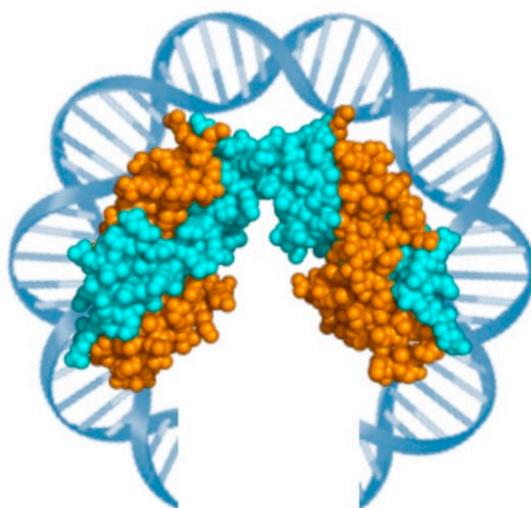
Table 1. Distribution, main structural features and activities of architectural proteins of hyperthermophilic archaea. “√” and “-” indicate the presence or absence, respectively, of a particular activity or feature.

Protein	HISTONE	ALBA	SUL7	CREN7	SMJ12
Archaeal sub-domain	Euryarchaea and Crenarchaea	Euryarchaea and Crenarchaea	Crenarchaea (<i>Sulfolobus</i>)	Crenarchaea	Crenarchaea (<i>S. solfataricus</i>)
Oligomeric Structure	Dimer (7.5 kDa)	Dimer (10 kDa)	Monomer (7 kDa)	Monomer (7 kDa)	Dimer (12 kDa)
DNA Binding	Cooperative	Cooperative	-	-	-
DNA Compaction		√	-	√	√
DNA Bending		√	-	√	√
DNA Modification		-	√	-	-
Supercoiling		Negative	Negative	Negative	Negative
Post-translational modifications	NO	Acetylation/Deacetylation	Methylation	Methylation	unknown

2.1. Histones

Archaea belonging to the sub-domains *Euryarchaea*, *Nanoarchaea* and *Thaumarchaea* and some *Crenarchaea* encode homologs of eukaryotic histones [11,14,17–21]. Archaeal histones possess a typical fold resembling eukaryotic H3 and H4, consisting of three hydrophobic α -helices, and interact with the DNA minor groove. In solution, archaeal histones form dimers, whereas they bind DNA as tetramers or, less frequently, as hexamers [21–23]. *In vitro*, each histone tetramer protects about 60 bases of dsDNA; at low histone concentrations, DNA is sharply bent in complexes, suggesting wrapping of DNA around a histone tetramer [18–21]. Under these conditions, histone tetramers induce negative supercoiling of circular DNA molecules, whereas positive supercoiling is observed at non-physiological ionic strength [19–21]. Many Archaea encode multiple histone homologs; the most studied histone proteins from hyperthermophilic archaea are HmfA and HmfB from *Methanothermobacter thermautotrophicus*. These proteins can form either homodimers or heterodimers, which differ in their DNA binding properties and compact DNA into nucleosome-like structures [18,22] (Figure 1). Histone paralogs can be differently expressed in different growth phases and conditions, suggesting that the dynamic histone composition may shape chromosome structure differently [24,25].

Figure 1. Structure of the *M. feravidus* histone tetramer in complex with DNA (PDB ID: 1B67), with the HmfA (orange) and HMfB (cyano) subunits.



Archaeal histones are considerably smaller than eukaryotic histones, due to the absence of the C- and N-terminal extensions that are targets of extensive post-translational regulatory modifications in eukaryotes [11]. No evidence for post-translational modification of archaeal histones has been observed, although protein acetyltransferase and methylase activities have been found in hyperthermophilic Archaea (see below).

Studies on eukaryotic chromatin established that nucleosomes are not positioned randomly in the genome, but rather different DNA segments facilitate nucleosome assembly depending on their primary sequence and the energy needed to wrap those fragments around the histones. In particular, alternating G/C- and A/T-rich dinucleotide tracts showed a propensity for histone-induced compaction; these observations led to the definition of a nucleosome positioning code [5]. Recent *in vitro* and *in vivo* experiments demonstrated that in the hyperthermophilic archaeal species *Methanothermobacter thermautotrophicus* and *Thermococcus kodakarensis* nucleosome assembly is directed by the same nucleosome positioning code observed in eukaryotes [26]. These results suggest that *in vivo* archaeal histones may use the same wrapping mechanisms as eukaryotic histones, although direct evidence is lacking. The chromatin organization of *T. kodakarensis* was also studied by applying a technology called chromatin particle spectrum analysis (CPSA), in which position and size of nucleosomal particles resistant to digestion by micrococcal nuclease were determined at the genomic level. This study demonstrated that *T. kodakarensis* chromatin particles consist of 30 bp units that can form linear multimers of variable length, up to ~450 bp. This structure is reminiscent of the so-called beads-on-a-string shape typical of eukaryotic chromatin; however, *T. kodakarensis* chromatin particles are in a dynamic equilibrium, in contrast to the static positioning of histones in eukaryotes. The 30-bp nucleosome units and their multimers were shown to colocalize with single or multiple, respectively, alternating G/C- A/T-rich dinucleotide tracts, a result consistent with the existence of a eukaryotic-type sequence preference code for nucleosome positioning in these organisms [27].

Another important issue is the relation between archaeal nucleosomes and gene expression. Several studies demonstrated that archaeal histones inhibit or reduce transcription by preventing preinitiation complex assembly and transcriptional initiation at promoters *in vitro*, suggesting that chromatin may play an important repressive function of basal archaeal gene expression *in vivo* [28,29]. Consistently, genome wide studies revealed that archaeal histones are excluded from genomic regions corresponding to transcription start sites, thus suggesting that, as in eukaryotes, promoters are nucleosome-free [26,27]. However, direct correlation between archaeal histone deposition and transcriptional status has not been established. Further studies are required to establish the exact structure of the nucleosomes in live archaeal cells and elucidate the relationships between nucleosomes and transcription. In particular, it would be interesting to assess whether the same chromatin structure and plasticity seen in *T. kodakarensis* are shared by all/other hyperthermophilic archaea; if this is the case, an attractive hypothesis is that chromatin plasticity might provide a mechanism to regulate gene expression by archaeal nucleosomes in the absence of the complex post-transcriptional control of eukaryotic histones.

2.2. Alba

Alba (Acetylation lowers binding affinity, reviewed in [30]) is a family of small, abundant DNA binding proteins (whose members are also known as Sac10 or Sso10, Ape10 *etc.*). These proteins are encoded by all thermophilic Archaea, most mesophilic Archaea and several eukaryotes [31–33]. In solution, Alba is a dimer of a 10-kDa subunit, which binds double-stranded DNA cooperatively without stringent sequence specificity and with high density (approximately 5 bp DNA per dimer), contacting the DNA minor groove; binding of Alba to DNA induces negative supercoiling, but not compaction [34–37].

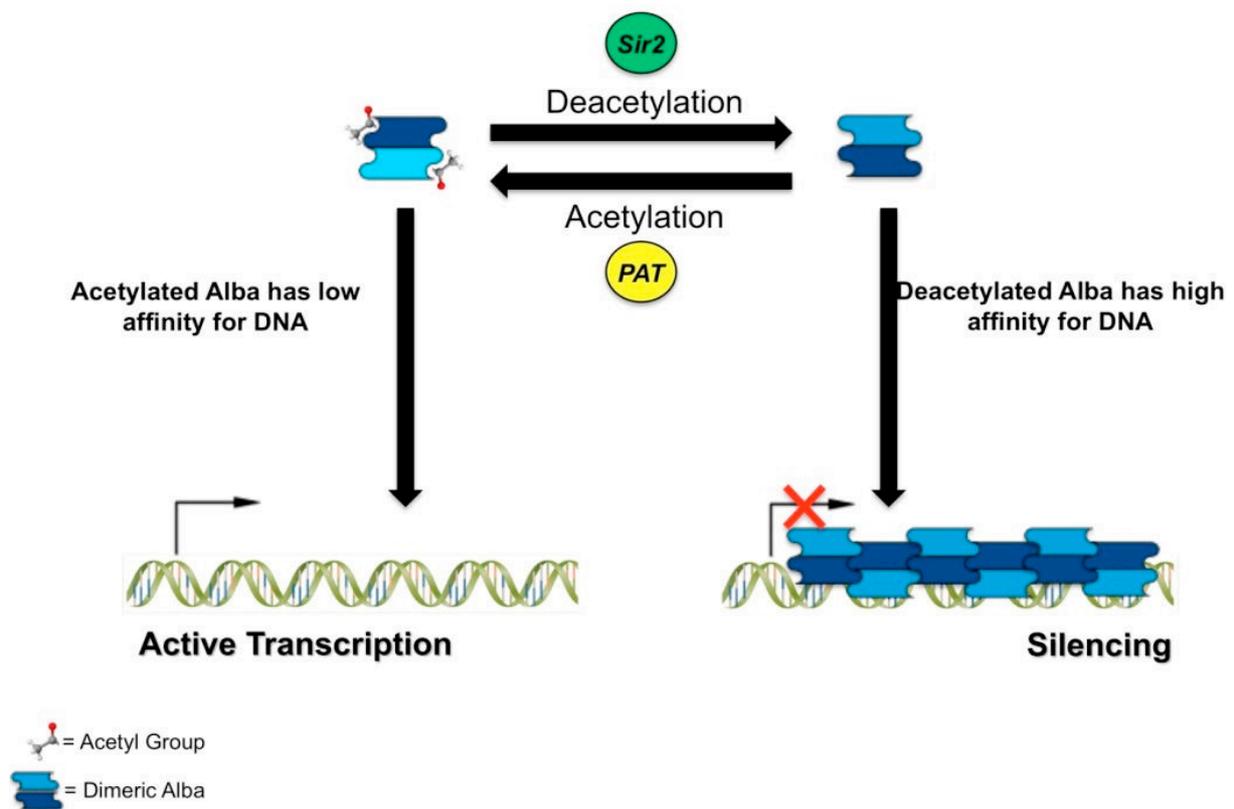
Electron microscopy studies revealed that the binding of *S. solfataricus* Alba to DNA forms extended interwound helical protein fibres [34,38]. Alba binding has two effects on DNA, depending on its concentration: at low protein:DNA ratio the protein is able to bridge two DNA molecules, while at higher concentrations Alba dimers bind cooperatively along DNA molecules, increasing their rigidity; dimer–dimer interactions promote the cooperative binding, but also appear to be responsible for bridging DNA molecules together [32,39]. Alba-DNA interaction has been studied using single-molecule tethered particle motion and optical tweezers, confirming that Alba binds cooperatively, inducing a 5-fold increase in the persistence length of the nucleoprotein filament. Moreover, Alba concentration-dependent dimer–dimer contacts between two nucleoprotein filaments were also observed [40].

Some Archaea encode multiple Alba paralogs. For instance, *Sulfolobus solfataricus* encodes two Alba proteins; the more abundant Alba1 and the so-called Alba2, which is only 5%–10% of the Alba1 amount and has lower affinity for DNA. Alba2 forms obligate heterodimers with Alba1 at physiological concentrations [38]. Alba2 lacks the F60 residue that is responsible for the cooperative binding of Alba1 dimers; consistently, whereas Alba1 yields rigid protein–DNA complexes, at similar protein:DNA ratios Alba1:Alba2 heterodimers form condensed protein–DNA complexes [39]. These results were confirmed by single-molecule techniques [41]. The dual binding mode of Alba and the existence of multiple Alba paralogs with different DNA binding properties suggest that these proteins are well suited to play an important role in modelling chromatin structure by regulating the equilibrium between stiff and interlinked DNA. Chromatin immunoprecipitation experiments showed that *S. solfataricus* Alba1 is widely distributed at many loci along the genome, thus supporting its role in chromosomal organization [42]. In particular, the ability of Alba to bridge DNA molecules suggests that it might participate in organizing the genome into higher order loops, as those found in bacteria, although the existence of such structures in archaeal genomes has not been proven to date.

A breakthrough in the archaeal chromatin field was the discovery that in *S. solfataricus* cells Alba1 interacts with the archaeal homolog of the eukaryotic silencing protein Sir2 and this interaction regulates Alba1 DNA binding affinity [31]. *In vivo*, a significant fraction of *S. solfataricus* Alba1 protein is found specifically acetylated at the lysine 16 residue and this modification is reversed

in vitro by the deacetylase activity of Sir2. Acetylation determines a significant decrease of the affinity of Alba1 for DNA. Using *in vitro* systems Alba1 was shown to have a repressive effect on transcription; interestingly, the efficiency of transcriptional silencing correlates with Alba1 modification: the acetylated form is about three-fold less effective than the Sir2-deacetylated form, which is consistent with the relative affinity for DNA of the two protein forms (Figure 2) [31].

Figure 2. Diagram showing a model for the post-translational regulation of Alba and its effect on transcription.



Alba1 lysine 16 acetylation is specifically catalysed by an acetyltransferase called PAT, which is conserved not only in archaea, but also in bacteria, although in this latter it seems to play a metabolic role [42,43]. Incubation of Alba1 with PAT reduces Alba1 affinity for DNA and this effect requires the presence of an acetyl-donor, thus confirming that PAT-mediated acetylation reduces Alba1 affinity for DNA. Although no data are available on the effect of Alba1 acetylation and Alba1-Sir2 interaction *in vivo*, an attractive model is that these proteins play a role not only in chromosome structure, but also in transcription regulation. Sir2-induced Alba deacetylation would stimulate recruitment of Alba to DNA (and possibly further deacetylation of other Alba molecules) resulting in spreading of a transcription repressive state due to Alba binding (Figure 2). Although in principle a similar model might be applied to other organisms encoding homologs of both Alba and Sir2, from archaea to higher eukaryotes, the general relevance of these findings is not yet clear. The functional interaction between

Alba and Sir2 and the effect of Alba on transcription is conserved the malaria protozoan *Plasmodium falciparum* [44]. However, the Lys-16 residue is not conserved within the Alba protein family, raising the possibility that in some members surrogate acetylated residues exist, which need to be identified to fully understand Alba's role in chromatin regulation.

The crystal and solution structures of Alba proteins from several archaeal species have been resolved [32,38,45–50]. These structures revealed that Alba shows a fold similar to that of the *N*-terminal domain of DNaseI and the *C*-terminal domain of bacterial translation factor IF3. The structural similarity to this latter factor, an RNA-binding protein, suggested that Alba may also have some RNA-related function, a hypothesis supported by the observation that Alba can bind to RNA *in vivo* and *in vitro* [51]. However, if this hypothesis is correct, such a function remains to be addressed. The 3D structure revealed that the protein forms a dimer of dimers, in which the dimer–dimer interface is stabilized by several hydrophobic residues centered around a phenylalanine (F60 in *S. solfataricus* Alba1) critical for dimerization [39,46]. Interestingly, the lysine residue corresponding to the *S. solfataricus* Alba1 K16, which is the target of the regulatory PAT-induced acetylation and Sir2-dependent deacetylation, is also involved in inter-dimer interaction, thus suggesting an elegant model linking the effect of Sir2-induced deacetylation to Alba oligomerization and DNA binding efficiency [38,39,46].

Resolution of an Alba2-DNA complex showed that each protein dimer contacts the minor groove and covers 4 DNA bases [49]. Binding induces a conformational rearrangement of the protein which facilitates protein oligomerization. The two types of dimer–dimer interactions were also observed in the DNA-bound structures; whereas interactions between adjacent dimers are mediated by chains of hydrogen bonds, the dimer–dimer interface between two DNA–protein fibers is stabilized mainly by interaction between conserved hydrophobic residues [49]. In all complexes, DNA was in its extended form, leaving the question of the role of Alba in DNA condensation still open. It is possible that *in vivo* the cooperation with other factors (such as histones, Sul7, Cren7, or analogous proteins) is needed to achieve appropriate compaction level. Thus, it would be interesting to analyse the physical and functional interaction of Alba with other chromatin proteins.

In eukaryotes chromatin activities, including its repressive effects on transcription, are modulated by post-translational histone modifications, such as acetylation, methylation and phosphorylation. Interestingly, the crystal structure of the Alba acetylase PAT in complex with coenzyme A revealed structure similarity to eukaryotic histone acetyltransferases, suggesting an intriguing analogy between Alba and eukaryotic histones [43]. However, in contrast to Alba, whose deacetylation induces transcriptional silencing, eukaryotic histone acetylation determines transcription activation, although the mechanism of acetylation-dependent transcription regulation is not completely clear. An intriguing hypothesis is that eukaryotic histones acetylation regulates transcriptional activity with a mechanism similar to Alba, although with opposite effects, *i.e.*, acetylation of histone tails may disrupt intermolecular interactions in higher order chromatin structures [46]. Confirmation of this model would require further studies on

archaeal as well as eukaryotic chromatin proteins.

2.3. *Sul7* and *Cren7*

Sul7 (formerly known as *Sso7d*, *Sac7d* or *Ssh7d*, reviewed in [1,13]) is a 7-kDa basic, abundant, non-specific DNA-binding protein found only in Crenarchaea of the genus *Sulfolobus*, where it accounts for 3%–5% of total protein content and binds strongly to double-stranded DNA without sequence preference. Binding protects DNA from thermal denaturation, elevating the melting point by 30 °C [52]. Moreover, it promotes the renaturation of complementary DNA strands at temperatures higher than the melting point of the duplex [53]. The annealing activity, which is strictly homology-dependent, might assist renaturation of the double helix at high temperature in processes requiring transient denaturation, such as transcription, recombination and repair. On the basis of these findings it can be suggested that *Sul7* plays a key role in stabilization of DNA at high temperature.

The 3-D structure of *Sul7*, solved both by NMR and X-ray crystallography [52,54–56] showed that the protein consists of two orthogonal anti-parallel β -sheets (one triple- and one double-stranded). This folding is reminiscent of that of the SH3 domain found in several eukaryotic proteins. Crystal structure of *Sul7*-DNA complex showed that it binds the DNA minor groove and induces changes in the helical twist and marked DNA bending (60°). These observations have been confirmed by *in vitro* functional assays, showing that *Sul7* induces bending of short DNA fragments and compaction of circular DNA molecules of any topology (negative, relaxed, or positive [57]).

Sul7 induces negative supercoiling of DNA in association with DNA topoisomerases. In particular *Sul7* and the *Sulfolobus* DNA topoisomerase TopoVI (see below) induce negative supercoiling of circular DNA molecules at physiological temperatures (up to 80 °C), transforming the conformational changes induced by *Sul7* into topological changes [57]. These results suggest that *Sul7* plays a role in DNA packaging and in the regulation of DNA superhelicity in *S. solfataricus*. Moreover, *Sul7* inhibits the positive supercoiling activity of reverse gyrase (see below), probably by stabilizing the double strand and inhibiting transient exposure of single strand regions required for reverse gyrase binding (see below). Indeed, *Sul7*-induced reverse gyrase inhibition is antagonized by the single strand binding protein, SSB, suggesting a functional interplay among *Sul7*, reverse gyrase and SSB in a physiological context [58].

Experiments *in vivo* showed that in *S. solfataricus* *Sul7* gene expression as well as the protein localization are affected by cell exposure to DNA damage: the *Sul7* gene transcription is repressed after cell exposure to UV light [59], whereas the protein dissociates from chromatin after treatment with alkylating agents [60].

Cren7 is also a small (about 7 kDa) monomeric, abundant chromatin protein conserved amongst hyperthermophilic crenarchaea [61]. It is similar in many respects to *Sul7*, although they are different

at primary sequence level. Indeed, Cren7 binds the minor groove of DNA non-specifically and the DNA interacting surface is a triple-stranded β -sheet. 3D structures of Cren7 and its complex with dsDNA showed that it shares the same SH3-like fold already found in Sul7. Upon binding, Cren7 binds compacts and kinks the dsDNA sharply, constrains negative DNA supercoils *in vitro* and is associated with genomic DNA *in vivo*. Molecular dynamics simulations at different temperatures indicate that Cren7 stabilizes the DNA duplex, while DNA molecules undergo B-like to A-like form transitions with increasing temperature [62].

For their functional and structural similarities, Sul7 and Cren7 could be both involved in genome packaging; although direct *in vivo* evidence is lacking, the fact that they are encoded by organisms (Crenarchaea) lacking histones makes this assumption likely. However, it should be noted that, whereas Cren 7 is present in all Crenarchaea, Sul7 is only found in the genus *Sulfolobus*, raising the question of the functional relationship and redundancy of the two proteins in different archaeal strains. In this respect, some significant differences in the structure of the two protein-DNA complexes and their *in vitro* activity have been found. Cren7 contains a large loop in the DNA binding surface, which is lacking in Sul7; this loop contains residues important for DNA binding of Cren7 [63,64], including one lysine residue, which undergoes reversible methylation [61] (see also below). Moreover, Cren7, but not Sul7, shows a *N*-terminal tail comprising serine and lysine residues; although evidence is lacking, an intriguing hypothesis is that these residues are targets of post-translational modifications, like in eukaryotic histones. The binding site of Cren7 is larger than that of Sul7 (8 vs. 4 bp), and biochemical experiments showed that Cren7 is twice as efficient as Sul7d in constraining negative supercoils [61], although this conclusion has been recently challenged by results obtained by combining atomic force microscopy and magnetic tweezers with molecular dynamics studies. These experiments demonstrated that the interaction of the two proteins with DNA is similar, as their binding affinity and extent of DNA compaction [65]. Moreover, Cren7- and Sul7-DNA complexes differ in flexibility from analogous bacterial and eukaryotic DNA-bending proteins [65].

Interestingly, both Sul7 and Cren7 are found to be monomethylated *in vivo* at specific lysine residues; for Cren7, these residues are located in the flexible loop close to the DNA-interacting surface [61], whereas the five Sul7 lysine residues found specifically methylated *in vivo* are not involved in DNA binding [52]. Consistently, lysine methylation affects the DNA binding affinity of Cren7 but not Sul7d [61]. Recent work identified a lysine methyltransferase called aKMT4 as a candidate factor responsible for this post-translational modification [66,67]. aKMT4 is well conserved in the genomes of Crenarchaea and shows striking structural and functional similarity to the eukaryotic histone methyltransferase KMT4/Dot1. Sul7 and Cren7 are both substrates of aKMT4-induced methylation *in vitro*. Methylation of Sul7, but not Cren7, is significantly stimulated by the presence of DNA; in particular, the level and efficiency of Sul7 methylation by aKMT4 are increased by pre-incubation of the protein with DNA. Since Sul7 methylation *in vivo* occurs only at lysine residues not

involved in DNA binding, the results of the *in vitro* experiments suggest that aKMT4-induced methylation might occur on the chromatin-bound Sul7, and a possible regulation of aKMT4 activity by the local chromatin environment [67]. Whereas methylation in the flexible loop might regulate the Cren7 DNA binding affinity in chromatin the functional significance of this modification remains to be elucidated for Sul7. The extent of Sul7 lysine methylation increases *in vivo* with increasing growth temperature, suggesting a heat-shock response related functional relevance [52]. These observations suggest that the Sul7 and Cren7-DNA interaction might be regulated differently, leading to the speculation that they may also have distinct functions; for instance, they may regulate dynamically chromosomal organization in response to different metabolic or environment conditions, or control the access of different proteins to chromatin. However, no data are available to support such hypotheses and further studies are required to address these points.

It is interesting to note that, in contrast to archaeal histones, three archaeal architectural proteins, Alba, Sul7 and Cren7 are all subject to post-translational modifications (acetylation or methylation). Extensive post-translational modifications play essential roles in establishing the epigenetic regulation of eucaryotic histones. Thus, post-translational modification of chromatin proteins represents an ancient and evolutionary conserved model for regulation of chromatin accessibility.

Sul7 was demonstrated to inhibit the activity of the DNA topoisomerase reverse gyrase [57] (see below); moreover, both Sul7 and Cren7 were shown to inhibit the DNA polymerase B1 from *S. solfataricus* in its strand displacement activity, which is likely involved in Okazaki fragment maturation during replication. Sul7 and Cren7 inhibit the polymerase ability to displace DNA–DNA, but not DNA–RNA hybrids, thus suggesting that the chromosomal proteins might take part in this process, directing the polymerase activity to removal of RNA primers while inhibiting extensive displacement of the newly synthesized DNA strand [68].

2.4. Other Architectural Proteins

Another protein affecting DNA conformation is Smj12 [69]. Smj12 is a member of the so-called BA (Bacterial-Archaeal) family, a large family of Helix-Turn-Helix DNA-binding proteins widespread in Archaea, and shares significant aminoacid identity with the transcriptional repressor Lrs14 [70]. Smj12 is a 12 kDa very basic protein, dimeric in solution and highly thermostable. Smj12 is a non-specific DNA binding protein that protects double-stranded DNA from thermal denaturation. Unlike Sul7 or Cren7, Smj12 does not compact DNA or induce bending *in vitro*, rather it induces positive supercoiling of DNA in topological assays. *In vivo* it is far less abundant than either Sul7 or Alba, suggesting that it is unlikely to organize higher order structures over the whole chromosome, but rather might have more specific functions [69].

CC1 is a small basic protein found only in a few Crenarchaea, whose function has not been

elucidated. CC1 is a 6-kDa, monomeric, basic protein that is expressed at a high level in *Thermoproteus tenax*. It resembles Sul7 and Cren7 for its β -sheet rich structural organization, although it does not share amino acid sequence similarity with these proteins. Moreover, in contrast to both Sul7 and Cren7, CC1 binds both ssDNA and dsDNA with comparable affinity, although binding to ssDNA is highly cooperative [71].

3. DNA Topoisomerases from Hyperthermophilic Archaea

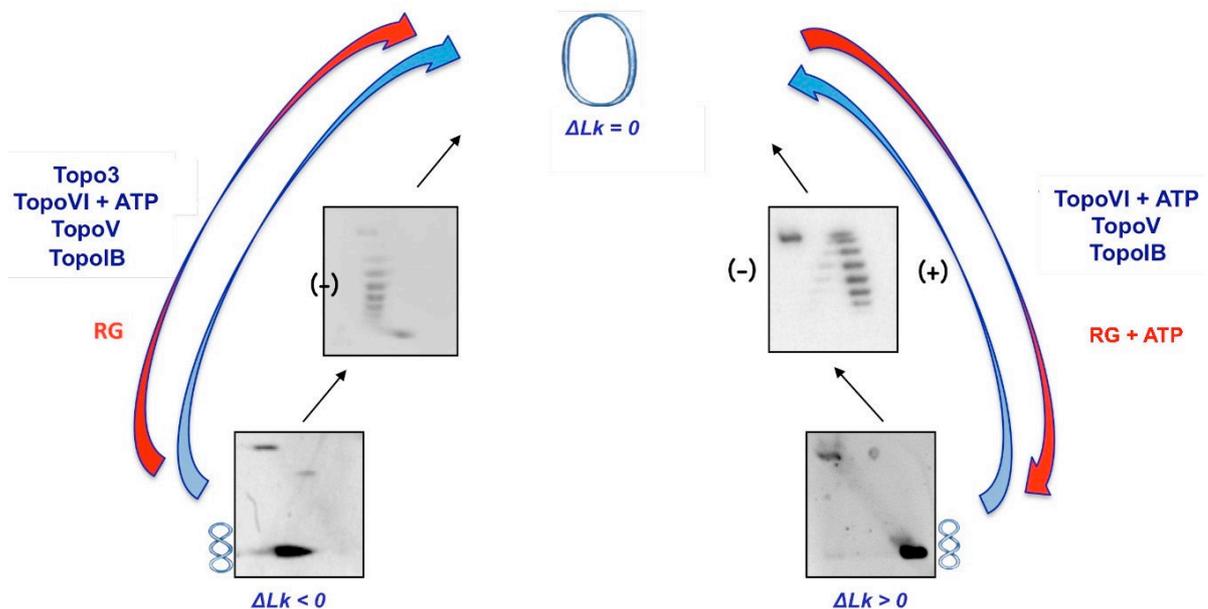
DNA topology is controlled and maintained by the action of DNA topoisomerases, essential enzymes that regulate the supercoiling level of DNA during all DNA activities (replication, transcription, recombination and repair). DNA supercoiling serves not only an important role in DNA compaction, but also regulates protein–DNA interactions and affects DNA transactions. DNA topoisomerases are classified according to their structure, specific activity and cellular function [2–4,10] (Table 2). DNA topoisomerases of Type I are monomeric, ATP-independent enzymes which induce a transient break in one DNA strand and pass the other intact strand through this “hole”, leading to DNA relaxation. The type I topoisomerase family comprises three sub-types, which differ in both structure and activity: Type IA topoisomerases relax only negative supercoils and are found in Bacteria, Eukaryotes and Archaea, whereas Type IB enzymes relax both negative and positive supercoils and are mostly eukaryotic. The type C sub-family comprises at the moment only one member from a hyperthermophilic archaeon (see below). Type II topoisomerases are heteromeric enzymes (dimer or tetramers), induce transient breaks of both DNA strands and relax both positive and negative supercoiling at the expense of ATP hydrolysis. Also for the TopoII family two sub-types (A and B) have been found so far, which differ significantly in their structure and action mechanism. Type IIA topoisomerases are ubiquitous in Bacteria and Eukarya, whereas members of the IIB family are mostly present in Archaea and plants.

Table 2. Features of DNA topoisomerases of hyperthermophilic archaea. Type IA enzymes are shared by almost all archaea, bacteria and eukaryotes; type IIB are present in all archaea and some plants; reverse gyrase is found only in hyperthermophilic archaea and bacteria; type IC is restricted to *M. kandleri*.

Enzyme	Topo VI	Topo 3	Reverse Gyrase	Topo V
Type	II B	IA	IA	IC
Structure	Heterotetramer A2 + B2 (A domain containing Winged Helix Domain (Active Site)/B domain containing ATP Binding Site)	Monomer (C- terminal domain implicate in DNA binding/ <i>N</i> -terminal domain implicate in topoisomerase activity)	Monomer (C- terminal domain like Topoisomerases type IA/ <i>N</i> -terminal domain like SF2 helicases)	Monomer (<i>N</i> -terminal domain with topoisomerase activity/ <i>C</i> - terminal domain with AP site processing activity)
	ATP dependent positive and negative supercoiled DNA relaxation; Cut double strand	ATP independent negative supercoiled DNA relaxation; Cut single strand	ATP dependent positive supercoiled DNA; ATP independent negative supercoiled DNA relaxation; Cut single strand	Positive and negative supercoiled DNA relaxation; Cut single strand
Activity				

The topological state of chromosomal DNA in hyperthermophilic archaea is not clear, although several lines of evidence suggest that it is relaxed or positively supercoiled [15,16] and thus likely more resistant to thermal denaturation as compared with the negatively supercoiled DNA of bacteria. This peculiar conformation results from the concerted activity and balance of a number of DNA topoisomerases of different families, including one type IA (Topoisomerase 3), one type IIB (Topoisomerase VI), and two peculiar chimeric enzymes, the type IC/AP lyase Topoisomerase V and the type IA/helicase reverse gyrase (Figure 3) [10].

Figure 3. Activities of DNA topoisomerases of hyperthermophilic archaea on circular plasmids. The gels show typical migration of plasmid topological isomers: ΔLk (Linking number) < 0 (negatively supercoiled); $\Delta Lk = 0$ (relaxed); $\Delta Lk > 0$ (positively supercoiled). Red arrows indicate the direction of activity of reverse gyrase (RG), which is always directed toward increasing linking number (from negative, to relaxed, to positive topoisomers). Blue arrows indicate the activity of all other topoisomerases, e.g., relaxation of either negative or positive topoisomers, with the exception of Topo3, which only relaxes negative supercoiling.



3.1. Topoisomerase VI

DNA Topoisomerase VI (TopoVI) [72,73] is classified as a Type IIB enzyme. Indeed, TopoVI is capable of cleaving both DNA strands, catalysing the passage of a DNA duplex

through the double-strand break and religating the product resulting in relaxation of either negative or positive topological stress. The reaction is strictly ATP-dependent. However, whereas TopoIIA cleavage leaves a four base overhang, there are only two after Topo VI cleavage [74]. TopoVI is a heterotetramer comprising two A and two B subunits, each encoded by a separate gene. Despite the similar organization, with the tyrosine responsible for DNA cleavage in the A subunit and the ATP binding site in the B subunit, TopoVI shares limited similarities with type IIA topoisomerases. As in TopoIIA enzymes, the TopoVI B subunit contains an ATP-binding site located within a protein domain known as the Bergerat fold, which is also found in other ATPases [73,75]. Structural analysis has shown that upon ATP binding and hydrolysis the monomeric B subunit undergoes dramatic conformational changes leading to dimerization [76]. In contrast, the TopoVI A subunit is completely different from the one of TopoIIA enzymes [77], whereas it is homologous to the eukaryotic meiosis-specific phosphodiesterase Spo11 [73]. Interestingly, TopoVI shares with TopoIIA enzymes the so-called Toprim domain, involved in magnesium binding; however, whereas in type IIA topoisomerases the Toprim domain is located in the B subunit, it is located in the A subunit of TopoVI. Lastly, both enzyme classes show a Winged Helix (WH) domain containing the active site tyrosine, although these regions are different in their primary sequence.

The two available complete structures of TopoVI showed the so-called “twin-gate” architecture, also found in TopoIIA enzymes, in which the ATPase domains of the two B subunits, located at one side of the heterotetramer, are able to coordinate DNA sliding through the DNA cleaving sites located at the opposite side of the molecule. This is possible due to conformational changes occurring during the reaction that couple nucleotide hydrolysis with strand passage [78,79]. Thus, despite the lack of conservation at the primary sequence level, TopoII A and B share significant structural similarity and reaction mechanism.

3.2. *Topoisomerase 3*

Archaeal Topoisomerase 3 (Topo3) is a canonical Type IA DNA topoisomerase. These enzymes are present in all organisms, with a few exceptions [80] and catalyze the relaxation of negatively, but not positively supercoiled substrates; this reaction does not require nucleotide hydrolysis. Experiments performed in many different organisms showed clearly that Topo3 enzymes play many roles in the cells, including regulation of the supercoiling level, and are involved in transcription, recombination and repair (see below). They are widespread in

hyperthermophilic archaea, although only a few members have been studied. Topo3 from *S. solfataricus* (SsTop3), consistent with its hyperthermophilic source, relaxes negative supercoiling and works optimally at 75 °C; however, ssDNA cleavage occurs even at lower temperatures (25–50 °C), whereas ligation of the cleaved DNA requires temperatures higher than 45 °C. In addition, SsTop3 induces efficient annealing of complementary ssDNA, an activity not shared by all Topo3 enzymes; annealing may participate in the catalytic cycle, stimulating religation of the DNA strand [81,82].

Deletion of Topo3 gene in *Sulfolobus islandicus* is not lethal, although the mutant growth rate is retarded with respect to the wild-type strain, in particular under nutrient deprivation conditions. The knock-out mutant shows alterations in the cell cycle, a high frequency of impaired DNA segregation and cell division and significant changes in the global transcriptional profile [83]. Taken together, these results suggest that archaeal Topo3 may be involved in chromosomal segregation and regulation gene expression through control of the level of intracellular DNA supercoiling.

More recently, a peculiar Topo3 enzyme from the endoparasitic hyperthermophile *Nanoarchaeum equitans* (NeqTop3) has been identified. In contrast to other topoisomerases of this family, NeqTop3 is a heterodimer consisting of two polypeptides encoded by a split gene. Besides the classical relaxation of negatively supercoiled DNA, this enzyme was shown to also catalyze a distinct reaction, the synthesis and dissolution of hemicatenanes. This reaction is due to DNA strand passage through denatured bubbles in the substrate DNA, which can be transiently exposed at the high temperature of incubation. At lower NeqTop3 concentrations, hemicatenanes are removed [84].

In both bacteria and eukaryotes, Topo3 enzymes interact physically and/or functionally with RecQ family helicases, and these complexes play multiple roles in recombination, repair, replication, and transcription. Analogously, SsTop3 was shown to interact with an archaeal RecQ-like helicase, Hel112. This enzyme shares limited sequence homology with eukaryotic RecQ helicases, but catalyses similar activities, including coordinate DNA unwinding and annealing, processing of synthetic Holliday junctions and stabilization of model replication forks [85,86]. SsTop3 inhibits the Hel112 helicase activity and stimulates formation and stabilization of Holliday junctions. The interplay between Hel112 and SsTop3 might regulate the equilibrium between recombination and anti-recombination activities at replication forks [86].

3.3. Reverse Gyrase

Reverse gyrase is an enzyme with peculiar structure and function. It consists of a Type IA topoisomerase module fused to a SF2 helicase-like domain, and induces ATP-dependent positive supercoiling of DNA; the reaction requires high temperature (>60 °C) [87–92]. Reverse gyrase is considered a thermophile-specific marker, since its gene is invariably present in the genomes of all bacteria and archaea living above 80 °C (and in some living at intermediate temperatures) but in no mesophilic organisms [93–95]. Positive DNA supercoiling increases the resistance of DNA to denaturation and consistently, plasmids isolated from hyperthermophiles show higher linking number as compared with plasmids from mesophiles [15,16]. Moreover, reverse gyrase protects DNA incubated at high temperature from depurination and degradation [96]. Based on these observations, a role for reverse gyrase in adaptation to high temperature has long been suggested; however, direct evidence of such a role is still lacking. Genetic experiments showed that deletion of the single reverse gyrase gene in the species *T. kodakarensis* is not lethal [97], although the mutant is less thermoresistant than the wild-type. In contrast, although *S. islandicus* encodes two reverse gyrase genes, neither could be deleted, suggesting that both are essential [98]. Certainly, convincing experiments in other species are required to ascertain the role of reverse gyrase in adaptation to high temperature.

The 3D structure of two reverse gyrases has been resolved, only one of which is from an archaeon, *Archaeoglobus fulgidus* [99]. Comparison of this structure with that of the enzyme from the bacterium *Thermotoga maritima* [100] revealed a well conserved type IA topoisomerase fold for the C-terminal domain, and a less conserved fold for the N-terminal domain, which contains a typical ATP-binding fold resembling that of the SF2 family helicases.

Mutational analysis of several archaeal reverse gyrases has identified motifs essential for DNA topoisomerase, ATPase and DNA binding activity [101–104] as well as the role of the so-called latch sub-domain, a region connecting the ATPase and topoisomerase modules [105,106].

Significant diversity in the details and optimal conditions of the reaction is seen among different archaeal reverse gyrases: their temperature optima range from 75 to 95 °C and ionic strength tolerance from 400 and up to 1200 mM; all RGs require a nucleotide (preferably ATP) for positive supercoiling reaction, however, in the absence of the nucleotide some, but not all, RGs are able to relax DNA; some enzymes are very processive, whereas other show a distributive behaviour of their positive supercoiling reaction [99,101–107]. The recently characterized reverse gyrase from *Pyrobaculum calidifontis* (PcalRG) [104] shows even remarkable thermostability, with significant activity even at 100 °C, and DNA binding and topoisomerase activity up to 1.2

M NaCl. The structural bases of these differences are not clear.

Given its chimeric helicase-topoisomerase structure, a long-standing question is whether there is any functional analogy between reverse gyrase and RecQ–Topo3 complexes (reviewed in [108]). The *in vitro* activity of such complexes from a number of mesophilic bacteria and eukaryotes have been characterized, showing that they can use coordinate DNA unwinding, annealing and topoisomerase activities to catalyse complex reactions, such as reversal of replication fork, branch migration and resolution of model recombination intermediates [109,110]. The structural similarity of reverse gyrase with RecQ–Topo3 complexes stimulated a wave of studies to establish whether they are also functionally similar. One set of experiments was aimed at testing whether RecQ–Topo3 complexes may show positive supercoiling activity: co-incubation of human RecQ1 and Topo3 α or *E. coli* RecQ and Topo3 failed to catalyse positive supercoiling at mesophilic temperature [111]; these observations, however, are of uncertain significance, since even reverse gyrase (the only topoisomerase shown to induce positive supercoiling) is not able to do so below 60 °C. Recently, the availability of thermophilic SsTop3 and Hel112 allowed the direct comparison of the activities of this complex with those of reverse gyrase. Although the SsTop3–Hel112 complex shares some activities with reverse gyrase on oligonucleotide substrates (see below), it was unable to induce positive supercoiling at high temperature under the same conditions that allow reverse gyrase to catalyse this reaction [86]. Thus, positive supercoiling is a peculiar activity of reverse gyrase and not a basic property of helicase-topoisomerase complexes.

The second set of experiments was aimed at testing whether reverse gyrase shows DNA unwinding, annealing and branch migration activities typical of RecQ and RecQ–Topo3 complexes. In earlier studies using classical helicase assays reverse gyrase failed to show processive DNA unwinding activity [101,102,112]. However, one of the two *S. solfataricus* reverse gyrases was shown to destabilize *in vitro* synthetic substrates resembling Holliday junctions (HJ) [113]. In this reaction, the enzyme does not act as a processive helicase; indeed, the reaction does not require the presence of ATP or a functional ATPase activity; moreover, mutational analysis showed that the topoisomerase activity is also dispensable. The enzyme binds at the core of the HJ, inducing a structural distortion that likely facilitates junction unfolding at high temperature.

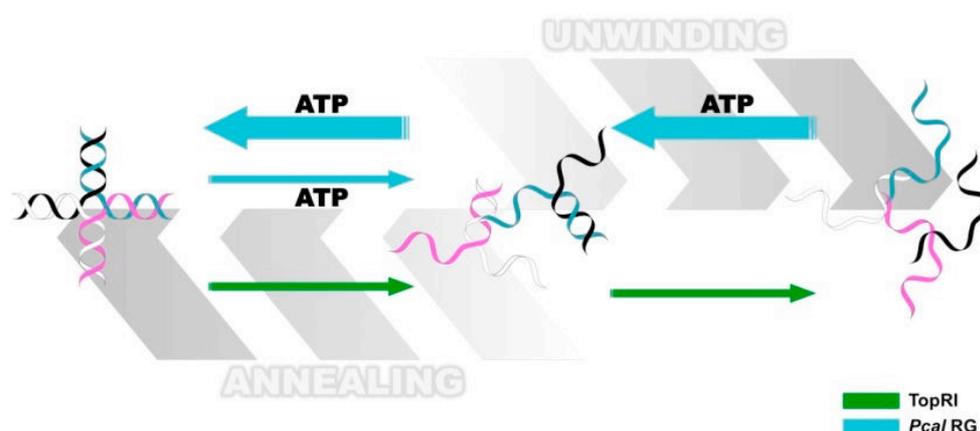
More recently, it was reported that PcalRG shows a real helicase activity, namely ATP-hydrolysis dependent unwinding of ds DNA and HJ structures [104]; interestingly, at higher enzyme/DNA ratios the reaction is reversed leading to re-annealing of DNA (Figure 4).

Although PcalRG is the only reverse gyrase for which processive ATP-dependent unwinding

and annealing activities have been demonstrated, it is possible that these abilities are shared by other reverse gyrases. Indeed, the reverse gyrase from the bacterium *T. maritima* was shown to induce transient unwinding of a short DNA duplex; the short-lived unwound intermediate does not accumulate because the reaction is rapidly reversed in the presence of ATP, and can only be seen using an ATP analog [114]. Combined unwinding and annealing activities could help clarify the mechanism of the positive supercoiling reaction. Indeed, unwinding of a DNA region in a topologically closed molecule would create two domains, one positive and one negative; a topoisomerase IA activity capable of relaxing only this latter would result in positive supercoiling of the final product [115,116].

The importance of DNA supercoiling modulation by reverse gyrase is confirmed by the involvement of the enzyme in the cell response to DNA damage and its interaction with chromatin and repair proteins [57,58,60,117,118]. In *S. solfataricus* reverse gyrase is recruited to chromatin after UV irradiation [96]; moreover, it forms a complex with the single-strand binding protein, SSB and the translesion DNA polymerase, PolY [118]. Interestingly, reverse gyrase inhibits PolY activity, and inhibition depends on the positive supercoiling activity of reverse gyrase. These results suggested the hypothesis that PolY might be sequestered onto highly positively supercoiled regions of the genome when its activity is not required, and illustrate an example of how the interaction with a chromatin protein might provide a means to control the potentially mutagenic activity of PolY under normal growth conditions.

Figure 4. Diagram comparing the activities of the *S. solfataricus* TopR1 and *P. calidifontis* PcalRG reverse gyrases on synthetic Holliday junctions. Light blue arrows indicate PcalRG activities (ATP-dependent unwinding and ATP-independent annealing); green arrows indicate TopR1 ATP-independent unwinding; the four DNA strands are shown in white, pink, black and blue, respectively. Whereas TopR1 unwinding activity proceeds up to single strand oligonucleotides, the main products of PcalRG unwinding are Y-shaped forks, likely because single strand products are annealed back [104,113].



Reverse gyrase also interacts with Sul7, and Sul7 inhibits reverse gyrase positive supercoiling activity [57]; since Sul7 induces negative supercoiling (see above) it is possible that the interplay between these proteins with antagonizing effect in the context of chromatin might contribute to control of the homeostasis of supercoiling level.

3.4. Topoisomerase V

Another peculiar archaeal topoisomerase is Topoisomerase V (TopoV), an enzyme so far found only in the species *Methanopyrus kandleri*. Although it belongs to the type I family of topoisomerases, TopoV is unrelated to either type IA or IB enzymes. Like eukaryotic Topo IB enzymes, it is able to relax both positively and negatively supercoiled DNA forming a 3'-link with DNA [119]; the enzyme is extremely thermophilic (being active up to 122 °C) [120]. TopoV is unique among all known topoisomerases in its structure combining a topoisomerase and a DNA repair domain. The DNA topoisomerase module is located at N-terminus of the protein, whereas the C-terminal domain exhibits an apurinic/apyrimidinic site-processing activity (AP lyase) [121]. This latter domain contains 24 helix-hairpin-helix repeats

which were shown to contribute to the high salt resistance and processivity of TopoV [121,122]. Based on these peculiarities, Forterre [123] assigned TopoV to a third family of type I DNA topoisomerases, named Topo IC; he also hypothesised that this enzyme, like many other orphan proteins, could have a viral origin.

The crystal structure of the 61 kDa *N*-terminal fragment of TopoV revealed no structural similarity to other topoisomerases. In particular, the structure of the active site region is unique, suggesting no conservation in the cleavage and religation mechanism. Moreover, the active site is not exposed on the molecule surface, suggesting the need for extensive conformational changes during the catalytic cycle [124]. By using real-time single-molecule and micromechanical experiments TopoV was shown to relax DNA using a constrained swiveling mechanism, similarly to type IB enzymes, and to perform multiple DNA relaxation cycles before discharging from DNA. Relaxation efficiency is enhanced by DNA supercoiling, but is reduced by the contacts between protein and DNA. Thus, TopoV relaxes DNA using a similar overall mechanism as type IB molecules, but in a completely different structural context [125]; an interesting line of future studies would be whether TopoV might also accomplish the same functions as the eukaryotic TopoIB. Recently, structural and biochemical studies demonstrated that an *N*-terminal 69-kDa fragment of TopoV is the minimal fragment with both topoisomerase and AP lyase activities and a putative Lys residue involved in the AP lyase activity was identified [126].

3.5. Topoisomerase IB

Topoisomerase IB (TopoIB) was thought for a long time to be a eukaryotic specific enzyme. However, recently a TopoIB-like gene was found in the sequenced genomes of two archaea of the phylum Thaumarchaeota. Phylogenetic analyses suggest that a TopoIB-like gene was present in the last common ancestor of Archaea and Eucarya [127]. The function of this archaeal TopoIB is currently unknown.

4. Conclusions

Whereas a large amount of data on architectural proteins and DNA topoisomerases of hyperthermophilic archaea has been accumulated over the last years, much work is still needed to delineate a complete picture of chromatin structure and regulation in these organisms. Since most information on this topic has been gained using *in vitro* assays, one of

the current main limitations in the field is the lack of *in vivo* data. This has been due in part to difficulties in the cultivation of some hyperthermophilic archaea species under laboratory conditions, and to technical setbacks that only recently allowed genetic manipulation of a few species, and still hamper it for the majority of these organisms. Hopefully, advances in genetic and cell biology techniques for hyperthermophilic archaea will help fill the gap in the near future. In addition, many questions that need to be addressed concern the relationships among different chromatin proteins, and in particular of members of the Alba family, which is present in all hyperthermophilic archaea, and other architectural proteins, namely histones in euryarchaea and Cren7 (and Sul7) in crenarchaea. In general, data on chromatin protein–protein interactions are scarce; a complete map of the interactions of chromatin proteins with each other and with other cellular proteins will help elucidate the role of chromatin in information processes and eventually lead to the identification of chromatin remodelling factors in hyperthermophilic archaea, a still uncovered area.

Another important issue is the role of reverse gyrase in cell resistance to high temperature; the hypothesis that the enzyme is an important factor in this adaptation is tantalizing given the restriction of reverse gyrase to hyperthermophiles, but experiments on different organisms are required. Moreover, still new functions for DNA topology and topoisomerases are emerging in eukaryotes, such as activation of DNA damage response [128]; it would be thus of great relevance to determine the exact role of each topoisomerase in cellular processes of hyperthermophilic archaea, their association with transcription/replication machinery and their role in maintenance of genome stability.

Acknowledgments

Work in the authors' laboratory is supported by FIRB-Futuro in Ricerca RBFR12OO1G_002 “Nematic”; Merit RBNE08YFN3; Ministero degli Affari Esteri (L.401/1990).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Luijsterburg, M.S.; White, M.F.; van Driel, R.; Dame, R.T. The major architects of chromatin: Architectural proteins in bacteria, archaea and eukaryotes. *Crit. Rev. Biochem.*

- Mol. Biol.* **2008**, *43*, 393–418.
2. Champoux, J.J. DNA topoisomerases: Structure, function, and mechanism. *Annu. Rev. Biochem.* **2001**, *70*, 369–413.
 3. Wang, J.C. Cellular roles of DNA topoisomerases: A molecular perspective. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 430–440.
 4. Chen, S.H.; Chan, N.L.; Hsieh, T.S. New mechanistic and functional insights into DNA topoisomerases. *Annu. Rev. Biochem.* **2013**, *82*, 139–170.
 5. Segal, E.; Fondufe-Mittendorf, Y.; Chen, L.; Thåström, A.; Field, Y.; Moore, I.K.; Wang, J.P.; Widom, J. A genomic code for nucleosome positioning. *Nature* **2006**, *442*, 772–778.
 6. Trifonov, E.N. Cracking the chromatin code: Precise rule of nucleosome positioning. *Phys. Life Rev.* **2011**, *8*, 39–50.
 7. Meyer, S.; Becker, N.B.; Syed, S.H.; Goutte-Gattat, D.; Shukla, M.S.; Hayes, J.J.; Angelov, D.; Bednar, J.; Dimitrov, S.; Everaers, R. From crystal and NMR structures, footprints and cryo-electron-micrographs to large and soft structures: Nanoscale modeling of the nucleosomal stem. *Nucleic Acids Res.* **2011**, *39*, 9139–9154.
 8. Kelly, T.K.; Liu, Y.; Lay, F.D.; Liang, G.; Berman, B.P.; Jones, P.A. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **2012**, *22*, 2497–2506.
 9. Xiao, B.; Freedman, B.S.; Miller, K.E.; Heald, R.; Marko, J.F. Histone H1 compacts DNA under force and during chromatin assembly. *Mol. Biol. Cell.* **2012**, *23*, 4864–4871.
 10. Forterre, P. Introduction and historical perspective. In *DNA Topoisomerases and Cancer*; Yves Pommier, Ed.; National Institutes of Health: Bethesda, MD, USA, 2012; pp. 1–52.
 11. Sandman, K.; Reeve, J.N. Archaeal chromatin proteins: Different structures but common function? *Curr. Opin. Microbiol.* **2005**, *8*, 656–661.
 12. Zhang, Z.; Guo, L.; Huang, L. Archaeal chromatin proteins. *China Life Sci.* **2012**, *55*, 377–385.
 13. Driessen R.P.; Dame, R.T. Nucleoid-associated proteins in Crenarchaea. *Biochem. Soc. Trans.* **2011**, *39*, 116–121.
 14. Reeve, J.N.; Bailey, K.A.; Li, W.T.; Marc, F.; Sandman, K.; Soares, D.J. Archaeal histones: Structures, stability and DNA binding. *Biochem. Soc. Trans.* **2004**, *32*, 227–230.
 15. Forterre, P.; Bergerat, A.; Lopez-Garcia, P. The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiol. Rev.* **1996**, *18*, 237–248.

16. López-García, P. DNA supercoiling and temperature adaptation: A clue to early diversification of life? *J. Mol. Evol.* **1999**, *49*, 439–452.
17. DeLange, R.J.; Green, G.R.; Searcy, D.G. A histone-like protein (HTa) from *Thermoplasma acidophilum*. I. Purification and properties. *J. Biol. Chem.* **1981**, *256*, 900–904.
18. Sandman, K.; Krzycki, J.A.; Dobrinski, B.; Lurz, R.; Reeve, J.N. HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 5788–5791.
19. Musgrave, D.R.; Sandman, K.M.; Reeve, J.N. DNA binding by the archaeal histone HMf results in positive supercoiling. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 10397–10401.
20. Musgrave, D.; Forterre, P.; Slesarev, A. Negative constrained DNA supercoiling in archaeal nucleosomes. *Mol. Microbiol.* **2000**, *35*, 341–349.
21. Marc, F.; Sandman, K.; Lurz, R.; Reeve, J.N. Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. *J. Biol. Chem.* **2002**, *277*, 30879–30886.
22. Starich, M.R.; Sandman, K.; Reeve, J.N.; Summers, M.F. NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. *J. Mol. Biol.* **1996**, *255*, 187–203.
23. Fahrner, R.L.; Cascio, D.; Lake, J.A.; Slesarev, A. An ancestral nuclear protein assembly: Crystal structure of the *Methanopyrus kandleri* histone. *Protein Sci.* **2001**, *10*, 2002–2007.
24. Sandman, K.; Grayling, R.A.; Dobrinski, B.; Lurz, R.; Reeve, J.N. Growth-phase-dependent synthesis of histones in the archaeon *Methanothermus fervidus*. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 12624–12628.
25. Dinger, M.E.; Baillie, G.J.; Musgrave, D.R. Growth phase-dependent expression and degradation of histones in the thermophilic archaeon *Thermococcus zilligii*. *Mol. Microbiol.* **2000**, *36*, 876–885.
26. Nalabothula, N.; Xi, L.; Bhattacharyya, S.; Widom, J.; Wang, J.P.; Reeve, J.N.; Santangelo, T.J.; Fondufe-Mittendorf, Y.N. Archaeal nucleosome positioning *in vivo* and *in vitro* is directed by primary sequence motifs. *BMC Genomics* **2013**, *14*, doi:10.1186/1471-2164-14-391.
27. Maruyama, H.; Harwood, J.C.; Moore, K.M.; Paszkiewicz, K.; Durley, S.C.; Fukushima, H.; Atomi, H.; Takeyasu, K.; Kent, N.A. An alternative beads-on-a-string chromatin architecture in *Thermococcus kodakarensis*. *EMBO Rep.* **2013**, *14*, 711–777.
28. Soares, D.; Dahlke, I.; Li, W.T.; Sandman, K.; Hethke, C.; Thomm, M.; Reeve, J.N.

- Archaeal histone stability, DNA binding, and transcription inhibition above 90 degrees C. *Extremophiles* **1998**, *2*, 75–81.
29. Xie, Y.; Reeve, J.N. Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *J. Bacteriol.* **2004**, *186*, 3492–3498.
 30. Xuan, J.; Feng, Y. The archaeal Sac10b protein family: Conserved proteins with divergent functions. *Curr. Protein Pept. Sci.* **2012**, *13*, 258–266.
 31. Bell, S.D.; Botting, C.H.; Wardleworth, B.N.; Jackson, S.P.; White, M.F. The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* **2002**, *296*, 148–151.
 32. Wardleworth, B.N.; Russell, R.J.; Bell, S.D.; Taylor, G.L.; White, M.F. Structure of Alba: An archaeal chromatin protein modulated by acetylation. *EMBO J.* **2002**, *21*, 4654–4662.
 33. Aravind, L.; Iyer, L.M.; Anantharaman, V. The two faces of Alba: The evolutionary connection between proteins participating in chromatin structure and RNA metabolism. *Genome Biol.* **2003**, *4*, R64:1–R64:9.
 34. Lurz, R.; Grote, M.; Dijk, J.; Reinhardt, R.; Dobrinski, B. Electron microscopic study of DNA complexes with proteins from the Archaeobacterium *Sulfolobus acidocaldarius*. *EMBO J.* **1986**, *5*, 3715–3721.
 35. Xue, H.; Guo, R.; Wen, Y.; Liu, D.; Huang, L. An abundant DNA binding protein from the hyperthermophilic archaeon *Sulfolobus shibatae* affects DNA supercoiling in a temperature-dependent fashion. *J. Bacteriol.* **2000**, *182*, 3929–3933.
 36. Chen, L.; Chen, L.R.; Zhou, X.E.; Wang, Y.; Kahsai, M.A.; Clark, A.T.; Edmondson, S.P.; Liu, Z.J.; Rose, J.P.; Wang, B.C.; *et al.* The hyperthermophile protein Sso10a is a dimer of winged helix DNA-binding domains linked by an antiparallel coiled coil rod. *J. Mol. Biol.* **2004**, *341*, 73–91.
 37. Edmondson, S.P.; Kahsai, M.A.; Gupta, R.; Shriver, J.W. Characterization of Sac10a, a hyperthermophile DNA-binding protein from *Sulfolobus acidocaldarius*. *Biochemistry* **2004**, *43*, 13026–13036.
 38. Jelinska, C.; Conroy, M.J.; Craven, C.J.; Hounslow, A.M.; Bullough, P.A.; Waltho, J.P.; Taylor, G.L.; White, M.F. Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. *Structure* **2005**, *13*, 963–971.
 39. Jelinska, C.; Petrovic-Stojanovska, B.; Ingledew, W.J.; White, M.F. Dimer–dimer stacking interactions are important for nucleic acid binding by the archaeal chromatin protein Alba. *Biochem. J.* **2010**, *427*, 49–55.

40. Lu, Y.W.; Huang, T.; Tsai, C.T.; Chang, Y.Y.; Li, H.W.; Hsu, C.H.; Fan, H.F. Using single-molecule approaches to study archaeal DNA-binding protein Alba1. *Biochemistry* **2013**, *52*, 7714–7722.
41. Laurens, N.; Driessen, R.P.; Heller, I.; Vorselen, D.; Noom, M.C.; Hol, F.J.; White, M.F.; Dame, R.T.; Wuite, G.J. Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat. Commun.* **2012**, *3*, doi:10.1038/ncomms2330.
42. Marsh, V.L.; Peak-Chew, S.Y.; Bell, S.D. Sir2 and the acetyltransferase, Pat, regulate the archaeal chromatin protein, Alba. *J. Biol. Chem.* **2005**, *280*, 21122–21128.
43. Brent, M.M.; Iwata, A.; Carten, J.; Zhao, K.; Marmorstein, R. Structure and biochemical characterization of protein acetyltransferase from *Sulfolobus solfataricus*. *J. Biol. Chem.* **2009**, *284*, 19412–19419.
44. Goyal, M.; Alam, A.; Iqbal, M.S.; Dey, S.; Bindu, S.; Pal, C.; Banerjee, A.; Chakrabarti, S.; Bandyopadhyay, U. Identification and molecular characterization of an Alba-family protein from human malaria parasite *Plasmodium falciparum*. *Nucleic Acids Res.* **2012**, *40*, 1174–1190.
45. Chou, C.C.; Lin, T.W.; Chen, C.Y.; Wang, A.H. Crystal structure of the hyperthermophilic archaeal DNA-binding protein Sso10b2 at a resolution of 1.85 Angstroms. *J. Bacteriol.* **2003**, *185*, 4066–4073.
46. Zhao, K.; Chai, X.; Marmorstein, R. Structure of a Sir2 substrate, Alba, reveals a mechanism for deacetylation-induced enhancement of DNA binding. *J. Biol. Chem.* **2003**, *278*, 26071–26077.
47. Biyani, K.; Kahsai, M.A.; Clark, A.T.; Armstrong, T.L.; Edmondson, S.P.; Shriver, J.W. Solution structure, stability, and nucleic acid binding of the hyperthermophile protein Sso10b2. *Biochemistry* **2005**, *44*, 14217–14230.
48. Kahsai, M.A.; Vogler, B.; Clark, A.T.; Edmondson, S.P.; Shriver, J.W. Solution structure, stability, and flexibility of Sso10a: A hyperthermophile coiled-coil DNA-binding protein. *Biochemistry* **2005**, *44*, 2822–2832.
49. Tanaka, T.; Padavattan, S.; Kumarevel, T. Crystal structure of archaeal chromatin protein Alba2-double-stranded DNA complex from *Aeropyrum pernix* K1. *J. Biol. Chem.* **2012**, *287*, 10394–10402.
50. Hada, K.; Nakashima, T.; Osawa, T.; Shimada, H.; Kakuta, Y.; Kimura, M. Crystal structure and functional analysis of an archaeal chromatin protein Alba from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. *Biosci. Biotechnol. Biochem.* **2008**, *72*, 749–758.

51. Guo, R.; Xue, H.; Huang, L. Ssh10b, a conserved thermophilic archaeal protein, binds RNA *in vivo*. *Mol. Microbiol.* **2003**, *50*, 1605–1615.
52. Baumann, H.; Knapp, S.; Lundbäck, T.; Ladenstein, R.; Härd, T. Solution structure and DNA-binding properties of a thermostable protein from the archaeon *Sulfolobus solfataricus*. *Nat. Struct. Biol.* **1994**, *1*, 808–819.
53. Guagliardi, A.; Napoli, A.; Rossi, M.; Ciaramella, M. Annealing of complementary DNA strands above the melting point of the duplex promoted by an archaeal protein. *J. Mol. Biol.* **1997**, *267*, 841–848.
54. Agback, P.; Baumann, H.; Knapp, S.; Ladenstein, R.; Härd, T. Architecture of nonspecific protein–DNA interactions in the Sso7d-DNA complex. *Nat. Struct. Biol.* **1998**, *5*, 579–584.
55. Edmondson, S.P.; Qiu, L.; Shriver, J.W. Solution structure of the DNA-binding protein Sac7d from the hyperthermophile *Sulfolobus acidocaldarius*. *Biochemistry* **1995**, *34*, 13289–13304.
56. Gao, Y.G.; Su, S.Y.; Robinson, H.; Padmanabhan, S.; Lim, L.; McCrary, B.S.; Edmondson, S.P. The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nat. Struct. Biol.* **1998**, *5*, 782–786.
57. Napoli, A.; Zivanovic, Y.; Bocs, C.; Buhler, C.; Rossi, M.; Forterre, P.; Ciaramella, M. DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of *Sulfolobus solfataricus*. *Nucleic Acids Res.* **2002**, *30*, 2656–2662.
58. Napoli, A.; Valenti, A.; Salerno, V.; Nadal, M.; Garnier, F.; Rossi, M.; Ciaramella, M. Functional interaction of reverse gyrase with single-strand binding protein of the archaeon *Sulfolobus*. *Nucleic Acids Res.* **2005**, *33*, 564–576.
59. Salerno, V.; Napoli, A.; White, M.F.; Rossi, M.; Ciaramella, M. Transcriptional response to DNA damage in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.* **2003**, *31*, 6127–6138.
60. Valenti, A.; Napoli, A.; Ferrara, M.C.; Nadal, M.; Rossi, M.; Ciaramella, M. Selective degradation of reverse gyrase and DNA fragmentation induced by alkylating agent in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.* **2006**, *34*, 2098–2108.
61. Guo, L.; Feng, Y.; Zhang, Z.; Yao, H.; Luo, Y.; Wang, J.; Huang, L. Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea. *Nucleic Acids Res.* **2008**, *36*, 1129–1137.
62. Chen, L.; Zhang, J.L.; Yu, L.Y.; Zheng, Q.C.; Chu, W.T.; Xue, Q.; Zhang, H.X.; Sun, C.C. Influence of hyperthermophilic protein Cren7 on the stability and conformation of

DNA: Insights from molecular dynamics simulation and free energy analysis. *J. Phys. Chem. B* **2012**, *116*, 12415–12425.

63. Feng, Y.; Yao, H.; Wang, J. Crystal structure of the crenarchaeal conserved chromatin protein Cren7 and double-stranded DNA complex. *Protein Sci.* **2010**, *19*, 1253–1257.
64. Zhang, Z.; Gong, Y.; Guo, L.; Jiang, T.; Huang, L. Structural insights into the interaction of the crenarchaeal chromatin protein Cren7 with DNA. *Mol. Microbiol.* **2010**, *76*, 749–759.
65. Driessen, R.P.; Meng, H.; Suresh, G.; Shahapure, R.; Lanzani, G.; Priyakumar, U.D.; White, M.F.; Schiessel, H.; van Noort, J.; Dame, R.T. Crenarchaeal chromatin proteins Cren7 and Sul7 compact DNA by inducing rigid bends. *Nucleic Acids Res.* **2013**, *41*, 196–205.
66. Chu, Y.; Zhang, Z.; Wang, Q.; Luo, Y.; Huang, L. Identification and characterization of a highly conserved crenarchaeal protein lysine methyltransferase with broad substrate specificity. *J. Bacteriol.* **2012**, *194*, 6917–6926.
67. Niu, Y.; Xia, Y.; Wang, S.; Li, J.; Niu, C.; Li, X.; Zhao, Y.; Xiong, H.; Li, Z.; Lou, H.; *et al.* A prototypic lysine methyltransferase 4 from archaea with degenerate sequence specificity methylates chromatin proteins Sul7d and Cren7 in different patterns. *J. Biol. Chem.* **2013**, *288*, 13728–13740.
68. Sun, F.; Huang, L. *Sulfolobus* chromatin proteins modulate strand displacement by DNA polymerase B1. *Nucleic Acids Res.* **2013**, *41*, 8182–8195.
69. Napoli, A.; Kvaratskelia, M.; White, M.F.; Rossi, M.; Ciaramella, M. A novel member of the bacterial-archaeal regulator family is a nonspecific DNA-binding protein and induces positive supercoiling. *J. Biol. Chem.* **2001**, *276*, 10745–10752.
70. Napoli, A.; van der Oost, J.; Sensen, C.W.; Charlebois, R.L.; Rossi, M.; Ciaramella, M. An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. *J. Bacteriol.* **1999**, *181*, 1474–1480.
71. Luo, X.; Schwarz-Linek, U.; Botting, C.H.; Hensel, R.; Siebers, B.; White, M.F. CC1, a novel crenarchaeal DNA binding protein. *J. Bacteriol.* **2007**, *189*, 403–409.
72. Bergerat, A.; Gadelle, D.; Forterre, P. Purification of a DNA topoisomerase II from the hyperthermophilic archaeon *Sulfolobus shibatae*. A thermostable enzyme with both bacterial and eucaryal features. *J. Biol. Chem.* **1994**, *269*, 27663–27669.
73. Bergerat, A.; de Massy, B.; Gadelle, D.; Varoutas, P.C.; Nicolas, A.; Forterre, P. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **1997**, *386*, 414–417.
74. Buhler, C.; Lebbink, J.H.; Bocs, C.; Ladenstein, R.; Forterre, P. DNA topoisomerase VI

- generates ATP-dependent double-strand breaks with two-nucleotide overhangs. *J. Biol. Chem.* **2001**, *276*, 37215–37222.
75. Dutta, R.; Inouye, M. GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem. Sci.* **2000**, *25*, 24–28.
 76. Corbett, K.D.; Berger, J.M. Structure of the topoisomerase VI-B subunit: Implications for type II topoisomerase mechanism and evolution. *EMBO J.* **2003**, *22*, 151–163.
 77. Nichols, M.D.; de Angelis, K.; Keck, J.L.; Berger, J.M. Structure and function of an archeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. *EMBO J.* **1999**, *18*, 6177–6188.
 78. Corbett, K.D.; Benedetti, P.; Berger, J.M. Holoenzyme assembly and ATP-mediated conformational dynamics of topoisomerase VI. *Nat. Struct. Mol. Biol.* **2007**, *14*, 611–619.
 79. Graille, M.; Cladière, L.; Durand, D.; Lecointe, F.; Gadelle, D.; Quevillon-Cheruel, S.; Vachette, P.; Forterre, P.; van Tilbeurgh, H. Crystal structure of an intact type II DNA topoisomerase: Insights into DNA transfer mechanisms. *Structure* **2008**, *16*, 360–370.
 80. Forterre, P.; Gribaldo, S.; Gadelle, D.; Serre, M.C. Origin and evolution of DNA topoisomerases. *Biochimie* **2007**, *89*, 427–446.
 81. Dai, P.; Wang, Y.; Ye, R.; Chen, L.; Huang, L. DNA topoisomerase III from the hyperthermophilic archaeon *Sulfolobus solfataricus* with specific DNA cleavage activity. *J. Bacteriol.* **2003**, *185*, 5500–5507.
 82. Chen, L.; Huang, L. Oligonucleotide cleavage and rejoining by topoisomerase III from the hyperthermophilic archaeon *Sulfolobus solfataricus*: Temperature dependence and strand annealing-promoted DNA religation. *Mol. Microbiol.* **2006**, *60*, 783–794.
 83. Li, X.; Guo, L.; Deng, L.; Feng, D.; Ren, Y.; Chu, Y.; She, Q.; Huang, L. Deletion of the topoisomerase III gene in the hyperthermophilic archaeon *Sulfolobus islandicus* results in slow growth and defects in cell cycle control. *J. Genet. Genomics* **2011**, *38*, 253–259.
 84. Lee, S.H.; Siaw, G.E.; Willcox, S.; Griffith, J.D.; Hsieh, T.S. Synthesis and dissolution of hemicatenanes by type IA DNA topoisomerases. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E3587–E3594.
 85. De Felice, M.; Aria, V.; Esposito, L.; de Falco, M.; Pucci, B.; Rossi, M.; Pisani, F.M. A novel DNA helicase with strand-annealing activity from the crenarchaeon *Sulfolobus solfataricus*. *Biochem. J.* **2007**, *408*, 87–95.
 86. Valenti, A.; de Felice, M.; Perugini, G.; Bizard, A.; Nadal, M.; Rossi, M.;

- Ciaramella, M. Synergic and opposing activities of thermophilic RecQ-like helicase and topoisomerase 3 proteins in Holliday junction processing and replication fork stabilization. *J. Biol. Chem.* **2012**, *287*, 30282–30295.
87. D’Amaro, A.; Rossi, M.; Ciaramella, M. Reverse gyrase: An unusual DNA manipulator of hyperthermophilic organisms. *Ital. J. Biochem.* **2007**, *56*, 103–109.
 88. Nadal, M. Reverse gyrase: An insight into the role of DNA-topoisomerases. *Biochimie* **2007**, *89*, 447–455.
 89. Perugino, G.; Valenti, A.; D’Amaro, A.; Rossi, M.; Ciaramella, M. Reverse gyrase and genome stability in hyperthermophilic organisms. *Biochem. Soc. Trans.* **2009**, *37*, 69–73.
 90. Valenti, A.; Perugino, G.; Rossi, M.; Ciaramella, M. Positive supercoiling in thermophiles and mesophiles: Of the good and evil. *Biochem. Soc. Trans.* **2011**, *39*, 58–63.
 91. Lulchev, P.; Klostermeier, D. Reverse gyrase-recent advances and current mechanistic understanding of positive DNA supercoiling. *Nucleic Acids Res.* **2014**, *42*, 8200–8213.
 92. Vettone, A.; Perugino, G.; Rossi, M.; Valenti, A.; Ciaramella, M. Genome stability: Recent insights in the topoisomerase reverse gyrase and thermophilic DNA-alkyltransferase. *Extremophiles* **2014**, *18*, 895–904.
 93. Forterre, P. A hot story from comparative genomics: Reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet.* **2002**, *18*, 236–237.
 94. Brochier-Armanet, C.; Forterre, P. Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* **2007**, *2*, 83–93.
 95. Heine, M.; Chandra, S.B. The linkage between reverse gyrase and hyperthermophiles: A review of their invariable association. *J. Microbiol.* **2009**, *47*, 229–234.
 96. Kampmann, M.; Stock, D. Reverse gyrase has heat-protective DNA chaperone activity independent of supercoiling. *Nucleic Acids Res.* **2004**, *32*, 3537–3545.
 97. Atomi, H.; Fukui, T.; Kanai, T.; Morikawa, M.; Imanaka, T. Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. *Archaea* **2004**, *1*, 263–267.
 98. Zhang, C.; Tian, B.; Li, S.; Ao, X.; Dalgaard, K.; Gökce, S.; Liang, Y.; She, Q. Genetic manipulation in *Sulfolobus islandicus* and functional analysis of DNA repair genes. *Biochem. Soc. Trans.* **2013**, *41*, 405–410.
 99. Rodríguez, A.C.; Stock, D. Crystal structure of reverse gyrase: Insights into the positive supercoiling of DNA. *EMBO J.* **2002**, *21*, 418–426.
 100. Rudolph, M.G.; del Toro Duany, Y.; Jungblut, S.P.; Ganguly, A.; Klostermeier, D.

- Crystal structures of *Thermotoga maritima* reverse gyrase: Inferences for the mechanism of positive DNA supercoiling. *Nucleic Acids Res.* **2013**, *41*, 1058–1070.
101. Déclais, A.C.; Marsault, J.; Confalonieri, F.; de La Tour, C.B.; Duguet, M. Reverse gyrase, the two domains intimately cooperate to promote positive supercoiling. *J. Biol. Chem.* **2000**, *275*, 19498–19504.
 102. Valenti, A.; Perugino, G.; D’Amaro, A.; Cacace, A.; Napoli, A.; Rossi, M.; Ciaramella, M. Dissection of reverse gyrase activities: Insight into the evolution of a thermostable molecular machine. *Nucleic Acids Res.* **2008**, *36*, 4587–4597.
 103. Li, J.; Liu, J.; Zhou, J.; Xiang, H. Functional evaluation of four putative DNA-binding regions in *Thermoanaerobacter tengcongensis* reverse gyrase. *Extremophiles* **2011**, *15*, 281–291.
 104. Jamroze, A.; Perugino, G.; Valenti, A.; Rashid, N.; Rossi, M.; Akhtar, M.; Ciaramella, M. The reverse gyrase from *Pyrobaculum calidifontis*, a novel extremely thermophilic DNA topoisomerase endowed with DNA unwinding and annealing activities. *J. Biol. Chem.* **2014**, *289*, 3231–3243.
 105. Rodríguez, A.C. Studies of a positive supercoiling machine. Nucleotide hydrolysis and a multifunctional “latch” in the mechanism of reverse gyrase. *J. Biol. Chem.* **2002**, *277*, 29865–29873.
 106. Rodríguez, A.C. Investigating the role of the latch in the positive supercoiling mechanism of reverse gyrase. *Biochemistry* **2003**, *42*, 5993–6004.
 107. Bizard, A.; Garnier, F.; Nadal, M. TopR2, the second reverse gyrase of *Sulfolobus solfataricus*, exhibits unusual properties. *J. Mol. Biol.* **2011**, *408*, 839–849.
 108. Larsen, N.B.; Hickson, I.D. RecQ helicases: Conserved guardians of genomic integrity. *Adv. Exp. Med. Biol.* **2013**, *767*, 161–184.
 109. Plank, J.L.; Wu, J.; Hsieh, T.S. Topoisomerase III α and Bloom’s helicase can resolve a mobile double Holliday junction substrate through convergent branch migration. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11118–11123.
 110. Bussen, W.; Raynard, S.; Busygina, V.; Singh, A.K.; Sung, P. Holliday junction processing activity of the BLM-Topo III α -BLAP75 complex. *J. Biol. Chem.* **2007**, *282*, 31484–31492.
 111. Valenti, A.; Ciaramella, M. National Research Council of Italy, Naples, Italy. Unpublished work, 2009.
 112. Capp, C.; Qian, Y.; Sage, H.; Huber, H.; Hsieh, T.S. Separate and combined biochemical

- activities of the subunits of a naturally split reverse gyrase. *J. Biol. Chem.* **2010**, *285*, 39637–39645.
113. Valenti, A.; Perugino, G.; Varriale, A.; D'Auria, S.; Rossi, M.; Ciaramella, M. The archaeal topoisomerase reverse gyrase is a helix-destabilizing protein that unwinds four-way DNA junctions. *J. Biol. Chem.* **2010**, *285*, 36532–36541.
 114. Ganguly, A.; del Toro Duany, Y.; Klostermeier, D. Reverse gyrase transiently unwinds double-stranded DNA in an ATP-dependent reaction. *J. Mol. Biol.* **2013**, *425*, 32–40.
 115. Jaxel, C.; Bouthier de la Tour, C.; Duguet, M.; Nadal, M. Reverse gyrase gene from *Sulfolobus shibatae* B12: Gene structure, transcription unit and comparative sequence analysis of the two domains. *Nucleic Acids Res.* **1996**, *24*, 4668–4675.
 116. Plank, J.; Hsieh, T.S. Helicase-appended topoisomerases: New insight into the mechanism of directional strand transfer. *J. Biol. Chem.* **2009**, *284*, 30737–30741.
 117. Napoli, A.; Valenti, A.; Salerno, V.; Nadal, M.; Garnier, F.; Rossi, M.; Ciaramella, M. Reverse gyrase recruitment to DNA after UV light irradiation in *Sulfolobus solfataricus*. *J. Biol. Chem.* **2004**, *279*, 33192–33198.
 118. Valenti, A.; Perugino, G.; Nohmi, T.; Rossi, M.; Ciaramella, M. Inhibition of translesion DNA polymerase by archaeal reverse gyrase. *Nucleic Acids Res.* **2009**, *37*, 4287–4295.
 119. Slesarev, A.I.; Stetter, K.O.; Lake, J.A.; Gellert, M.; Krah, R.; Kozyavkin, S.A. DNA topoisomerase V is a relative of eukaryotic topoisomerase I from a hyperthermophilic prokaryote. *Nature* **1993**, *364*, 735–737.
 120. Kozyavkin, S.A.; Pushkin, A.V.; Eiserling, F.A.; Stetter, K.O.; Lake, J.A.; Slesarev, A.I. DNA enzymology above 100 degrees C. Topoisomerase V unlinks circular DNA at 80–122 degrees C. *J. Biol. Chem.* **1995**, *270*, 13593–13595.
 121. Belova, G.I.; Prasad, R.; Nazimov, I.V.; Wilson, S.H.; Slesarev, A.I. The domain organization and properties of individual domains of DNA topoisomerase V, a type 1B topoisomerase with DNA repair activities. *J. Biol. Chem.* **2002**, *277*, 4959–4965.
 122. Pavlov, A.R.; Belova, G.I.; Kozyavkin, S.A.; Slesarev, A.I. Helix-hairpin-helix motifs confer salt resistance and processivity on chimeric DNA polymerases. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13510–13515.
 123. Forterre, P. DNA topoisomerase V: A new fold of mysterious origin. *Trends Biotechnol.* **2006**, *24*, 245–247.
 124. Taneja, B.; Patel, A.; Slesarev, A.; Mondragón, A. Structure of the N-terminal fragment of topoisomerase V reveals a new family of topoisomerases. *EMBO J.* **2006**, *25*, 398–408.

125. Taneja, B.; Schnurr, B.; Slesarev, A.; Marko, J.F.; Mondragón, A. Topoisomerase V relaxes supercoiled DNA by a constrained swiveling mechanism. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14670–14675.
126. Rajan, R.; Prasad, R.; Taneja, B.; Wilson, S.H.; Mondragón, A. Identification of one of the apurinic/aprimidinic lyase active sites of topoisomerase V by structural and functional studies. *Nucleic Acids Res.* **2013**, *41*, 657–666.
127. Brochier-Armanet, C.; Gribaldo, S.; Forterre, P. A DNA topoisomerase IB in Thaumarchaeota testifies for the presence of this enzyme in the last common ancestor of Archaea and Eucarya. *Biol. Direct.* **2008**, *3*, doi:10.1186/1745-6150-3-54.
128. Kumar, A.; Mazzanti, M.; Mistrik, M.; Kosar, M.; Beznoussenko, G.V.; Mironov, A.A.; Garrè, M.; Parazzoli, D.; Shivashankar, G.V.; Scita, G.; *et al.* ATR mediates a checkpoint at the nuclear envelope in response to mechanical stress. *Cell* **2014**, *158*, 633–646.

3. Conclusions and Future Perspectives

3.1 DNA Topology and Transcription

The process of transcription involves extensive interactions between RNA polymerase (RNAP) and a double-stranded DNA template. Each stage of transcription, and each reaction carried out by RNAP during transcription, is affected by the sequence context of the DNA template (Larson et al. 2011; Wang and Greene 2011; Dangkulwanich et al. 2014; Ruff et al. 2015; Washburn and Gottesman 2015). Therefore a quantitative understanding of how DNA sequence influences transcription output requires comprehensive knowledge of RNAP activity in all sequence contexts.

In Vvedenskaya et al., we report the development of a new experimental platform (called MASTER) that facilitates the analysis of the relationship between nucleic acid sequence and functional output during transcription for *Escherichia coli* RNAP on a scale previously inaccessible.

In particular, I analyzed the effects of DNA topology on the selection of transcription starting site (TSS) and I found that DNA topology is a determinant of this selection. I demonstrated that negatively supercoiled DNA provides a driving force for an expansion of transcription-bubble favoring the selection of a TSS at positions shifted downstream respect to the element -10 of *E. coli* promoter compared to non supercoiled DNA molecules (linear or relaxed). My analysis also reveals TSS-region sequence determinants that confer sensitivity or resistance to topology-dependent changes in TSS selection for a consensus core promoter.

MASTER can be readily adapted for comprehensive analysis of sequence determinants for transcription elongation, transcription termination, or any other biological process that involves complex nucleic-acid interactions.

Future studies could explore the effects of different topologies on transcription starting site selection; in particular it could be interesting evaluate the effect of positively supercoiled molecules as well as of partially negative/positive supercoiled DNA. Indeed plasmids from thermophiles and hyperthermophiles that contain reverse gyrase are generally relaxed or slightly positive (with superhelical densities of -0.015 to 0.013), while plasmids from mesophilic organisms that lack a reverse gyrase are negatively supercoiled (have superhelical densities of -0.048 to -0.068) (Lulchev and Klostermeier 2014). Therefore the analysis of the impact of positive supercoiling on transcription initiation is fundamental to investigate the determinants of transcription starting site selection in thermophilic organisms. Moreover in *E. coli* DNA supercoiling levels are under a tight homeostatic control (Menzel et al. 1983; Neumann and Quinones 1997; Snoep et al. 2002) and the overall level of negative superhelicity continuously decreases from exponential to stationary growth phase (Balke et al. 1987). So the analysis of DNA molecules with different negative supercoiling degrees could help to investigate about the impact of DNA topology in different phases of growth.

It has been shown for individual promoters in several bacterial species that transcription-coupled DNA supercoiling (TCDS) affects the transcription of neighboring genes (Richet and Raibaud 1991; Burns and Minchin 1994; Chen and Lilley 1999; Naughton et al. 2013). Recently Sobetzko analyzed the effect of TCDS on gene expression and chromosome architecture. He shows that the supercoiling sensitivity is one of the determinants of the local gene expression driving the arrangement of genes on the *E. coli* chromosome and he provides evidence that this rationale is conserved in other bacteria (Sobetzko 2016). Therefore it could be of great interest to analyze how positive supercoiling can affect the gene expression and chromosome architecture.

An in-deep analysis of the impact of DNA topology on DNA transcription is essential to elucidate the process of transcription in eukaryotes. Indeed it's known that in eukaryotes, the most potent generator of DNA supercoils is transcription by RNA polymerase. Over-wound DNA generated ahead of polymerase can change the twist of the DNA to give a tighter DNA helix, whereas under-wound DNA changes the twist to give a looser DNA helix (Corless and Gilbert 2016). DNA supercoiling also influences the translocation of polymerase between the strands of the DNA double helix. Under-wound DNA is more efficiently transcribed (Ma et al. 2013); conversely, over-wound DNA has a lower transcription rate and, at high levels, prevents the processivity of the polymerase complex because the tight DNA helix prevents DNA strand separation.

Finally to further understand how DNA supercoiling influences protein–DNA interaction, chromatin structure and gene regulation it is necessary to determine the properties of promoters and regulatory elements that are sensitive to DNA supercoiling. In this context MASTER offers the possibility to analyze up to potentially at least 4^{10} DNA sequences, allowing the analysis of the behavior of all possible sequence variants in a region of interest in different conditions *in vitro* and *in vivo*.

3.2 Structural diversity of DNA

Since its discovery, supercoiled DNA has been extensively studied by biologists and by physicists. The former were interested in its implication in biological processes, while the latter investigated its structure and studied its topology and conformation. DNA supercoiling plays key roles in genetic processes in the cell such as replication, transcription, recombination as well as DNA packaging (Sinden 1994). A lot of studies was performed in order to clarify the structural properties of supercoiled DNA and large amounts of experimental data are available in literature. In the last decade, Atomic Force Microscopy (AFM) has proven to be one of most valuable techniques for the studies of the biological structure and dynamic processes of DNA. Despite its great potential, all

studies performed until now used as substrate linear or negatively supercoiled DNA molecules while no studies on positively supercoiled DNA molecules have been performed by AFM.

In my work I analyzed, for the first time the positively supercoiled DNA in comparison with the negatively one in order to investigate structural peculiarities of both topologies. Moreover I analyzed both conformations also by classical biochemical techniques.

For the first the biochemical analyses revealed the presence of exposed bases when the DNA was negatively supercoiled but not in positively supercoiled DNA, supporting their biological roles. As already known, the exposed bases create local helix distortions that allow DNA to assume conformations that would be energetically unfavourable if the bases remained paired. The presence of single strand regions in negatively supercoiled DNA is in line with its biological role during replication and transcription. Conversely the absence of exposed bases in positively supercoiled molecules suggests that the positive conformation could play a key role in DNA stability maintenance preventing the local denaturation and protecting the DNA structure.

AFM data showed different conformations for negatively and positively DNA in almost all the conditions used. In particular I observed that in presence of a low concentration of divalent cations (3 mM $MgCl_2$), negatively supercoiled DNA molecules (with a superhelical density of -0.04) tend to adopt loosely interwound conformations (as already reported in Schmatko et al. 2014) while positively supercoiled DNA (with a superhelical density of +0.05) showed a typical plectonemic conformation. It could be hypothesized that the different electrostatic repulsion at crossings causes a different immobilization of supercoiled molecules on mica substrate: in the case of negatively supercoiled plasmids the repulsion exceeds bending and twisting costs. Therefore, a DNA molecule would prefer to be under high twist rather than having crossings. In the case of positively supercoiled, the electrostatic repulsion at crossings is probably reduced because of the different orientation of supercoiling and as a consequence we observe plectonemes. These results represent a starting point for further studies to identify possible differences between negative and positive supercoiling and to elucidate their biological roles *in vivo*.

Determining how the compacted genetic material can still remain accessible to replication/transcription machinery is of great importance for understanding the basic principles of genetic regulation. For this purpose AFM analyses in liquid represent a tool with enormous potential to investigate the accessibility and binding of proteins to both negatively and positively supercoiled DNA molecules. This technique permits nanometric-scale imaging of single molecules under near-physiological conditions and this advantage can be also used to study the dynamics of different biological processes involving negative and positive supercoiling (Alonso-Sarduy et al. 2011).

Recently it was reported that negatively circular supercoiled DNA could dynamically fold in particular higher-order structures called hyperplectonemes. Hyperplectonemic coiling could provide substantial compaction. This is especially relevant for prokaryotes, in which plectonemic coiling is thought to be the major mode of DNA packaging. The presence of different molecules in living organisms that have DNA condensing function in combination with high level of DNA negative supercoiling, suggests that hyperplectonemic compaction could be found *in vivo*. These studies illustrate how the bacterial nucleoid can be effectively compacted and organized, while remaining dynamic in nature and readily accessible to DNA binding proteins and processing enzymes (Japaridze et al. 2017).

Hyperthermophilic organisms require specific mechanisms inducing positive supercoiling to counteract the denaturing effect of their growth temperature (Valenti et al. 2011). Indeed, several lines of evidence suggest that DNA is more positive in (hyper)thermophilic archaea as compared with mesophiles: several plasmids from different hyperthermophilic strains were found in relaxed to positively supercoiled form (Forterre et al. 1996; Charbonnier and Forterre 1994). In addition, all hyperthermophiles invariably possess the DNA topoisomerase reverse gyrase, which is the only hyperthermophilic specific able to introduce positive supercoils in DNA (Brochier-Armanet and Forterre 2006). In addition to reverse gyrase, it is reported the presence in hyperthermophiles of DNA-binding proteins able to induce both positive and negative supercoiling (Napoli et al. 2001; Xue et al. 2000; Bell et al. 2002). These evidences are bringing out a complex picture concerning the genome topology of these organisms. In addition, (hyper)thermophilic archaea belonging to the subdomain Euryarchaeota contain tetrameric eukaryotic-like histones, which wrap DNA into positive supercoils at high salt concentrations, but into negative supercoils in low salt conditions (Musgrave 1991, 2000).

In this context the possibility to perform AFM analyses on positively supercoiled DNA at the larger scales (several tens of kilo base pairs) can give elucidation on the complex genomic situation in hyperthermophiles and can highlight the link between the nanomechanical properties of nucleoprotein complexes and the protein regulatory function as already seen for negatively supercoiled DNA.

3.3 Understanding Reverse Gyrase function

Reverse gyrase (RG) is a peculiar thermophilic DNA topoisomerase with the unique capability to introduce positive supercoils in DNA molecules. The physiological roles of reverse gyrase are not yet fully clear, although positive supercoiling is expected to protect DNA from denaturation at the growth temperatures of hyperthermophiles (Lulchev and Klostermeier 2014). The overwinding

activity has been demonstrated only *in vitro*. It is unclear whether positive DNA supercoiling is its *in vivo* function. Moreover a number of observations support a role of reverse gyrase in DNA protection and repair (Hsieh and Plank 2006; Kampmann and Stock 2004; Valenti et al. 2006, 2009; Napoli et al. 2004).

One of the most important approaches to study protein function *in vivo* is the labeling with a small polypeptide, which enable the *in vivo* detection of the resultant fusion protein. This approach offers also the advantage to simplify and optimize the purification procedures. In the last decades the green fluorescent protein (GFP) has been applied in numerous fields of cellular and molecular biology (Chalfie et al. 1994; Tsien 1998). However GFP tag and its derivatives have some limitations: (1) their relative big dimensions may affect the target protein's function and localization; (2) the isolation of the natural fluorophore in the active site makes GFPs and variants partially sensitive to environmental changes; (3) because the formation of the internal fluorophore is O₂-dependent, GFPs are not suitable tags in applications requiring anaerobic conditions; (4) the general use of most protein-tags is restricted to mesophiles and mild reaction conditions. Thermostable GFPs variants have been developed as tags for thermophilic microorganisms (Aliye et al. 2015; Cava et al. 2008; Pédelacq et al. 2006) but they suffer from most of the same limitations listed above.

A valid alternative to GFPs, is represented by SNAP-tagTM technology based on the fusion of O⁶-alkylguanine-DNA-alkyl-transferases (AGTs or MGMTs) to the protein of interest. These proteins can react *in vitro* and/or *in vivo* with benzyl-guanine (BG)-derivative substrates, leading to the labeling of a fused protein of interest with a desired chemical group. Recently, the discovery of a BG-sensitive thermostable OGT from the thermoacidophilic archaeon *S. solfataricus* (*SsOGT-H⁵*) opened the possibility to widen the SNAP-tagTM technology to organisms, which thrive in extreme environments. In order to investigate the biological role of RG *in vivo*, I used *SsOGT-H⁵* in fusion with *S. solfataricus* reverse gyrase (RG) in the archaeal model system *Sulfolobus islandicus*. The analysis of the H⁵-RG fusion protein demonstrated that it is possible to study the RG under its own physiological conditions, without the need to remove the tag. This new fusion protein gives the opportunity for further *in vivo* and *in vitro* studies.

To our knowledge is the first time that reverse gyrase is detected *in vivo* under physiological conditions. This discovery offers the possibility to an in-deep analysis of reverse gyrase function *in vivo* through experiment of protein-protein interaction or in situ localization. It's known that reverse gyrase is degraded after treatment of *S. solfataricus* with alkylating agent, parallel to the degradation of genomic DNA (Valenti et al. 2006). The possibility of *in vivo* labeling of RG could

be elucidated this aspect allowing to detect RG degradation directly in the cells under different conditions of treatment or by comparing wild-type and ogt-mutant strains.

Recently Couturier et al. described the first study reporting the specific detection of two reverse gyrases within *Sulfolobus solfataricus* P2 strain. TopR2 exhibits significant positive supercoiling activity at 45°C, a temperature at which TopR1 is not active. TopR1 is inessential at low temperature but required at high temperature and therefore probably involved in thermoadaptation and/or in DNA transaction processes during active division of *S. solfataricus* cells at 80°C. The SNAP-tag technology offers the advantage to label a protein with an infinite number of chemical groups (attached to BG substrate) and this possibility could be exploited to study the roles of the two reverse gyrases *in vivo* within the same organism.

Finally the high substrate specificity toward BG-derivatives makes H⁵ an interesting starting point to be modified by molecular evolution in order to obtain a thermostable variant active on Benzyl-Cytosine (BC)-derivative substrates, like the commercial CLIP-tagTM: this orthogonal substrate specificity will allow simultaneously and specifically labeling of different molecular targets in living cells (Gautier et al. 2008).

3.4 DNA Topoisomerases and their poisoning

Topoisomerases are ubiquitous enzymes involved in maintaining genomic stability of the cell by modulating the level of supercoiling of DNA strands. Besides their customary functions, topoisomerases are important cellular targets of widely used anticancer drugs. In particular, topoisomerase II α (Top2 α) has been postulated as the primary molecular target of anthracycline's anticancer activity (Mordente et al. 2017). Despite their wide use in chemotherapy, amthracyclines have a limiting side effect. In particular the cardiotoxicity is the most frequent adverse effect of both conventional and modern anticancer-targeted therapy, representing the leading noncancer-related cause of morbidity and mortality in long-term survivors (Hamo et al. 2015).

Coupling the drug to peptide carrier could improve solubility, decrease side effects and can offer the possibility of cell or tissue specific introduction (Hudecz et al. 1992; Duncan et al. 1987). Another approach for intracellular delivery is using cell-penetrating peptides like oligoarginines with cationic character; this method is less specific but could be highly efficient (Nakase et al. 2008; Futaki et al. 2005).

In my work, I analyzed the biochemical properties of a new oligoarginine-daunomycin conjugate previously described by Miklàn et al. (2009). It's known that anticancer activity of Daunomycin is based on both DNA binding properties and inhibition of human topoisomerase II α . So I investigated the inhibition effect of Arg-derivative on both human topoisomerase II α and topoisomerase I and I

found that it is able to inhibit both topoisomerases relaxation activity efficiently as Daunomycin. Then I analyzed the DNA binding properties of Daunomycin conjugate and I demonstrated that it is able to interact with DNA mainly by intercalative mode as already known for the free drug. Furthermore Arg-derivative is also able to stabilize the double helical structure of DNA as demonstrated by melting experiment. These results suggest that oligoarginine-daunomycin conjugate could be a promising alternative to Daunomycin in antitumor therapy.

References

1. Aliye N, Fabbretti A, Lupidi G, Tsekoa T, Spurio R. Engineering color variants of green fluorescent protein (GFP) for thermostability, pH-sensitivity, and improved folding kinetics. *Appl Microbiol Biotechnol* 2015, 99:1205–1216.
2. Balke V.L and Gralla J.D. Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J. Bacteriol.* 1987; 169, 4499–4506.
3. Bell S.D, Botting C.H, Wardleworth B.N, Jackson S.P, White M.F. The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* 2002; 29, 148–151.
4. Brochier-Armanet C and Forterre P. Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggest a complex history of vertical inheritance and lateral gene transfers. *Archaea*. 2007 May;2(2):83-93.
5. Burns H. and Minchin S. (1994) Thermal energy requirement for strand separation during transcription initiation: the effect of supercoiling and extended protein DNA contacts. *Nucleic Acids Res.* 1994; 22, 3840–3845.
6. Capranico G, Tinelli S, Austin C.A, Fisher M.L, Zunino F. Different patterns of gene expression of topoisomerase II isoforms in differentiated tissues during murine development. *Biochim. Biophys. Acta*, 1992, 1132, 43-48.
7. Cava F, de Pedro M.A, Blas-Galindo E, Waldo G.S, Westblade L.F, Berenguer J. Expression and use of superfolder green fluorescent protein at high temperatures in vivo: a tool to study extreme thermophile biology. *Environ Microbiol* 2008, 10:605–613.
8. Chalfie M, Tu Y, Euskirchen G, Ward W.W, Prasher D.C. Green fluorescent protein as a marker for gene expression. *Science* 1994, 263:802–805.
9. Charbonnier, F and Forterre P. Comparison of plasmid DNA topology among mesophilic and thermophilic eubacteria and archaeobacteria. *J. Bacteriol.* 1994; 176, 1251–1259.
10. Chen D and Lilley D.M. Transcription-induced hypersupercoiling of plasmid DNA. *J. Mol. Biol.* 1999; 285, 443–448.
11. Chen Y, Saari J.T, Kang Y.J. Weak antioxidant defenses make the heart a target for damage in copper-deficient rats. *Free Radic. Biol. Med.*, 1994, 17, 529-536.
12. Corless S and Gilbert N. Effects of DNA supercoiling on chromatin architecture. *Biophys Rev* (2016) 8:245–258 DOI 10.1007/s12551-016-0210-1.
13. Couturier M, Bizard A.H, Garnier F, Nadal M. Insight into the cellular involvement of the two Reverse gyrases from the hyperthermophilic Archaeon *Sulfolobus solfataricus*. *BMC Molecular Biology* 2014, 15:18.

14. Dangkulwanich M, Ishibashi T, Bintu L, Bustamante C. Molecular mechanisms of transcription through single-molecule experiments. *Chem Rev.* 2014 Mar 26;114(6):3203-23. doi: 10.1021/cr400730x.
15. Doroshow J.H and Davies K.J. Redox cycling of anthracyclines by cardiac mitochondria. II. Formation of superoxide anion, hydrogen peroxide, and hydroxyl radical. *J. Biol. Chem.*, 1986, 261, 3068-3074.
16. Doroshow J.H, Locker G.Y, Myers C.E. Enzymatic defenses of the mouse heart against reactive oxygen metabolites: alterations produced by doxorubicin. *J. Clin. Invest.*, 1980, 65, 128-135.
17. Duncan R, Kopečková-Rejmanova P, Strohalm J, Hume I, Cable H.C, Pohl J, Lloyd J.B, Kopecek J. Anticancer agents coupled to N-(2-hydroxypropyl)methacrylamide copolymers. I. Evaluation of daunomycin and puromycin conjugates in vitro. *Br. J. Cancer* 55 (1987) 165–174.
18. Force T and Wang Y. Mechanism-based engineering against anthracycline cardiotoxicity. *Circulation*, 2013, 128, 98-100.
19. Forterre P, Bergerat A, Lopez-Garcia P. The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiol. Rev.* 1996; 18, 237–248.
20. Futaki S, Membrane-permeable arginine-rich peptides and the translocation mechanisms. *Adv. Drug Deliv. Rev.* 57 (2005) 547–558.
21. Gammella E, Maccarinelli F, Buratti P, Recalcati S, Cairo G. The role of iron in anthracycline cardiotoxicity. *Front. Pharmacol.*, 2014, 5, 25.
22. Gautier A, Juillerat A, Heinis C, Corrêa IR Jr, Kindermann M, Beaufils F, Johnsson K. An engineered protein-tag for multiprotein labeling in living cells. *Chem Biol* 2008, 15:128–136.
23. Goffart S, von Kleist-Retzow J.C, Wiesner R.J. Regulation of mitochondrial proliferation in the heart: power-plant failure contributes to cardiac failure in hypertrophy. *Cardiovasc. Res.*, 2004, 64, 198-207.
24. Goormaghtigh E, Brasseur R, Huart P, Ruyschaert J.M. Study of the adriamycin-cardiolipin complex structure using attenuated total reflection infrared spectroscopy. *Biochemistry*, 1987, 26, 1789-1794.
25. Hamo C.E and Bloom M.W. Getting to the Heart of the Matter: An Overview of Cardiac Toxicity Related to Cancer Therapy. *Clin Med Insights Cardiol.* 2015; 9(Suppl 2): 47–51.

26. Hsieh T.S and Plank J.L. Reverse gyrase functions as a DNA renaturase: annealing of complementary single-stranded circles and positive supercoiling of a bubble substrate. *J. Biol. Chem.* 2006, 281, 5640–5647.
27. Hudecz F, Clegg J.A, Kajtá J, Embleton M.J, Szekerke M, Baldwin R.W. Synthesis, conformation, biodistribution, and in vitro cytotoxicity of daunomycin branched polypeptide conjugates. *Bioconjug. Chem.* 3 (1992) 49–57.
28. Irobalieva R.N, Fogg J.M, Catanese D.J, Sutthibutpong T, Chen M, Barker A.K, Ludtke S.J, Harris S.A, Schmid M.F, Chiu W, Zechiedrich L. Structural diversity of supercoiled DNA. *Nature Communications* 2015 DOI:10.1038/ncomms9440.
29. Japaridze A, Muskhelishvili G, Benedetti F, Gavriilidou A.F.M, Zenobi R, De Los Rios P, Longo G, Dietler G. Hyperplectonemes: A Higher Order Compact and Dynamic DNA Self Organization. DOI: 10.1021/acs.nanolett.6b05294 *Nano Lett* 2017.
30. Jiang Y and Yin Y. Analyzing DNA Structure Quantitatively at a Single-Molecule Level by Atomic Force Microscopy. *Atomic Force Microscopy Investigations into Biology - From Cell to Protein*, Dr. Christopher Frewin (Ed.), InTech, DOI: 10.5772/35796 2012.
31. Kampmann M and Stock D. Reverse gyrase has heat-protective DNA chaperone activity independent of supercoiling. *Nucleic Acids Res.* 2004, 32, 3537–3545.
32. Larson M.H, Landick R, Block S.M. Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. *Mol Cell.* 2011 Feb 4;41(3):249-62. doi:10.1016/j.molcel.2011.01.008.
33. Lulchev P and Klostermeier D. Reverse gyrase—recent advances and current Mechanistic understanding of positive DNA Supercoiling. 8200–8213 *Nucleic Acids Research*, 2014, Vol. 42, No. 13 Doi: 10.1093/nar/gku589.
34. Lyu Y.L, Kerrigan J.E, Lin C.P, Azarova A.M, Tsai Y.C, Ban Y, Liu L.F. Topoisomerase IIbeta mediated DNA double-strand breaks: implications in doxorubicin cardiotoxicity and prevention by dexrazoxane. *Cancer Res.*, 2007, 67, 8839-8846.
35. Lyubchenko Y.L and Shlyakhtenko L.S. Visualization of supercoiled DNA with atomic force Microscopy in situ. *Proc. Natl. Acad. Sci. USA* Vol. 94, pp. 496–501, January 1997.
36. Ma J, Bai L, Wang M.D. Transcription under torsion. *Science* 2013, 340: 1580–1583.
37. Menna P, Gonzalez Paz O, Chello M, Covino E, Salvatorelli E, Minotti G. Anthracycline cardiotoxicity. *Expert Opin. Drug Saf.*, 2011.
38. Menzel R and Gellert M. Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell.* 1983 Aug;34(1):105-13.

39. Mikla Z, Orba E, Csik G, Schlosser G, Magyar A, Hudecz F. New Daunomycin–Oligoarginine Conjugates: Synthesis, Characterization, and Effect on Human Leukemia and Human Hepatoma Cells. Published online 11 June 2009 in Wiley interscience (www.interscience.wiley.com). DOI 10.1002/bip.21264.
40. Minotti G, Cairo G, Monti E. Role of iron in anthracycline cardiotoxicity: new tunes for an old song? *FASEB J.*, 1999, 13, 199-212.
41. Minotti G, Menna P, Salvatorelli E, Cairo G, Gianni L. Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. *Pharmacol. Rev.*, 2004, 56, 185-229.
42. Mordente A, Meucci E, Martorana G.E, Giardina B, Minotti G. Human heart cytosolic reductases and anthracycline cardiotoxicity. *IUBMB Life*, 2001, 52, 83-88.
43. Mordente A, Meucci E, Silvestrini A, Martorana G.E, Giardina B. New developments in anthracycline-induced cardiotoxicity. *Curr. Med. Chem.*, 2009, 16, 1656-1672.
44. Mordentea A, Meuccia E, Martoranaa G.E, Taviana D, Silvestrinia A. Topoisomerases and Anthracyclines: Recent Advances and Perspectives In Anticancer Therapy and Prevention of Cardiotoxicity. *Current Medicinal Chemistry*, 2017, 24, 1-20.
45. Musgrave D, Forterre P, Slesarev A. Negative constrained DNA supercoiling in archaeal nucleosomes. *Mol. Microbiol.* 2000; 35, 341–349.
46. Musgrave D, Sandman K.M, Reeve J.N. DNA binding by the archaeal histone Hmf results in positive supercoiling. *Proc Natl Acad Sci U S A.* 1991 Dec 1;88(23):10397-401.
47. Nakase I, Takeuchi T, Tanaka G, Futaki S. Methodological and cellular aspects that govern the internalization mechanisms of arginine-rich cell-penetrating peptides. *Adv. Drug Deliv. Rev.* 60 (2008) 598–607.
48. Napoli A, Kvaratskeli M, White M.F, Rossi M, Ciaramella M. A novel member of the bacterial–archaeal regulator family is a nonspecific DNA-binding protein and induces positive supercoiling. *J. Biol. Chem.* 2001; 276, 10745–10752.
49. Napoli A, Valenti A, Salerno V, Nadal M, Garnier F, Rossi M, Ciaramella M. (2004) Reverse gyrase recruitment to DNA after UV light irradiation in *Sulfolobus solfataricus*. *J. Biol. Chem.* 2004, 279, 33192–33198.
50. Naughton C, Corless S, Gilbert N. Divergent RNA transcription: a role in promoter unwinding? *Transcription*, 2013, 4, 162–166.
51. Neumann, S. and Quinones A. Discoordinate gene expression of *gyrA* and *gyrB* in response to DNA gyrase inhibition in *Escherichia coli*. *J. Basic Microbiol.* 1997; 37, 53–69.

52. Pédelacq J.D, Cabantous S, Tran T, Terwilliger T.C, Waldo G.S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotech* 2006, 24:79–88.
53. Richet E and Raibaud O. Supercoiling is essential for the formation and stability of the initiation complex at the divergent malEp and malKp promoters. *J. Mol. Biol.* 1991; 218, 529–542.
54. Ruff E.F, Record M.T Jr, Artsimovitch I. Initial events in bacterial transcription initiation. *Biomolecules.* 2015 May 27;5(2):1035-62. doi: 10.3390/biom5021035.
55. Salvatorelli E, Guarnieri S, Menna P, Liberi G, Calafiore A.M, Mariggio M.A, Mordente A, Gianni L, Minotti G. Defective one- or two-electron reduction of the anticancer anthracycline epirubicin in human heart. Relative importance of vesicular sequestration and impaired efficiency of electron addition. *J. Biol. Chem.*, 2006, 281, 10990-11001.
56. Sarduy L.A, Roduit C, Dietler G, Kasas S. Human topoisomerase II–DNA interaction study by using atomic force microscopy. *FEBS Letters* 585 (2011) 3139–3145.
57. Schmatko T, Muller P, Maaloum M. Surface charge effects on the 2D conformation of Supercoiled DNA. *Soft Matter*, 2014, 10, 2520–2529.
58. Sinden R.R. *DNA Structure and Function*. Academic Press, San Diego, 1994, pp. 95–128.
59. Snoep J.L, van der Weijden, C.C, Andersen H.W, Westerhoff H.V, Jensen P.R. DNA supercoiling in *Escherichia coli* is under tight and subtle homeostatic control, involving gene-expression and metabolic regulation of both topoisomerase I and DNA gyrase. *Eur. J. Biochem.*, 2002; 269, 1662–1669.
60. Sobetzko P. Transcription-coupled DNA supercoiling dictates the chromosomal arrangement of bacterial genes. 1514–1524 *Nucleic Acids Research*, 2016, Vol. 44, No. 4 Published online 17 January 2016 doi: 10.1093/nar/gkw007.
61. Sterba M, Popelova O, Vavrova A, Jirkovsky E, Kovarikova P, Gersl V, Simunek T. Oxidative stress, redox signaling, and metal chelation in anthracycline cardiotoxicity and pharmacological cardioprotection. *Antioxid. Redox Signal.*, 2013, 18, 899-929.
62. Szabó R, Bánóczy Z, Mező G, Láng O, Kőhidai L, Hudecz F. Daunomycin-polypeptide conjugates with antitumor activity. *Biochimica et Biophysica Acta* 1798 (2010) 2209–2216.
63. Tokarska-Schlattner M, Zaugg M, Zuppinger C, Wallimann T, Schlattner U. New insights into doxorubicin-induced cardiotoxicity: the critical role of cellular energetics. *J. Mol. Cell. Cardiol.*, 2006, 41, 389-405.
64. Tsien R.Y. The green fluorescent protein. *Annu Rev Biochem* 1998, 67:509–544.

65. Valenti A, Napoli A, Ferrara M.C, Nadal M, Rossi M, Ciaramella M. Selective degradation of reverse gyrase and DNA fragmentation induced by alkylating agent in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.* 2006, 34, 2098–2108.
66. Valenti A, Perugino G, Nohmi T, Rossi M, Ciaramella M. Inhibition of translesion DNA polymerase by archaeal reverse gyrase. *Nucleic Acids Res.* 2009, 37, 4287–4295.
67. Valenti A, Perugino G, Rossi M, Ciaramella M. Positive supercoiling in thermophiles and Mesophiles: of the good and evil. *Biochemical Society Transactions* (2011) Volume 39, part 1.
68. Vejpongsa P and Yeh E.T. Topoisomerase 2 β : a promising molecular target for primary prevention of anthracycline-induced cardiotoxicity. *Clin. Pharmacol. Ther.*, 2014, 95, 45-52.
69. Vettone A, Serpe M, Hidalgo A, Berenguer J, del Monaco G, Valenti A, Rossi M, Ciaramella M, Perugino G. A novel thermostable protein tag: optimization of the *Sulfolobus Solfataricus* DNA-alkyl-transferase by protein engineering. *Extremophiles* (2016) 20:1–13 DOI 10.1007/s00792-015-0791-9.
70. Vvedenskaya I.O, Zhang Y, Goldman S-R, Valenti A, Visone V, Taylor D.M, Ebright R.H, Nickels B.E. Massively Systematic Transcript End Readout, "MASTER": Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. *Mol Cell.* 2015 Dec 17;60(6):953-65. doi: 10.1016/j.molcel.2015.10.029.
71. Wang F and Greene EC. Single-molecule studies of transcription: from one RNA polymerase at a time to the gene expression profile of a cell. *J Mol Biol.* 2011 Oct 7;412(5):814-31. doi 10.1016/j.jmb.2011.01.024.
72. Washburn R.S and Gottesman M.E. Regulation of transcription elongation and termination. *Biomolecules.* 2015 May 29;5(2):1063-78. doi: 10.3390/biom5021063.
73. Xue H, Guo R, Wen Y, Liu D, Huang L. An abundant DNA binding protein from the hyperthermophilic archaeon *Sulfolobus shibatae* affects DNA supercoiling in a temperature dependent fashion. *J. Bacteriol.* 2000; 182, 3929–3933.
74. Zhang S, Liu X, Bawa-Khalfe T, Lu L.S, Lyu Y.L, Liu L.F, Yeh E.T. Identification of the molecular basis of doxorubicin-induced cardiotoxicity. *Nat. Med.*, 2012, 18, 1639-1642.