# Phylogenomics of type II DNA topoisomerases

Danièle Gadelle, Jonathan Filée, Cyril Buhler, and Patrick Forterre*

## Summary

**Type II DNA topoisomerases (Topo II) are essential enzymes implicated in key nuclear processes. The recent discovery of a novel kind of Topo II (DNA topoisomerase VI) in Archaea led to a division of these enzymes into two non-homologous families, (Topo IIA and Topo IIB) and to the identification of the eukaryotic protein that initiates meiotic recombination, Spo11. In the present report, we have updated the distribution of all Topo II in the three domains of life by a phylogenomic approach. Both families exhibit an atypical distribution by comparison with other informational proteins, with predominance of Topo IIA in Bacteria, Eukarya and viruses, and Topo IIB in Archaea. However, plants and some Archaea contain Topo II from both families. We confront this atypical distribution with current hypotheses on the evolution of the three domains of life and origin of DNA genomes.**
***BioEssays* 25:232–242, 2003.**
**© 2003 Wiley Periodicals, Inc.**

## Introduction

DNA molecules in all cellular organisms are subjected to topological constraints. These constraints arise from the size and the bi-helical structure of the DNA. In the double helix, the two strands are topologically linked in vivo (as long as the phosphodiester backbone is intact) either by ring closure (in the case of circular DNA) or by the binding of macromolecular complexes. This topological structure prevents the free rotation of the two strands around each other, unless they can cross each other via transient breaks in the phosphodiester backbone. DNA topoisomerases (EC 5.99.1.3) appear to have evolved to solve this topological problem and participate in all DNA transactions that require partial or complete unwinding of the two DNA strands (replication, transcription, recombination, chromatin remodelling) (for reviews, see Refs.

1,2). In addition, specific DNA topoisomerases can introduce supercoils in closed circular DNA duplexes, determining both DNA secondary and tertiary structures via changes in the twist and the writhing of DNA duplexes.

In order to modify the number of topological links (linking number, Lk) between the two DNA strands, DNA topoisomerases catalyse the crossing of single- or double-strand DNA molecules via the formation of transient single- or double-strand breaks. These breaks are characterised by a phosphotyrosine bond either at the 3′ end or the 5′ end of the broken DNA strands. This covalent linkage allows DNA topoisomerases to store the energy of the cleavage reaction that will be used in the following ligation step.

DNA topoisomerases have attracted much attention in the last two decades as intracellular targets of major antibacterial and antitumoral drugs (for a recent review see Ref. 3). Most of these drugs act by interfering with the processing of the DNA break, causing stabilisation of the transient covalent DNA–DNA topoisomerase complex into a form that is harmful for the cell. The therapeutic importance of DNA topoisomerases has been dramatically emphasised recently with the systematic use of the DNA gyrase inhibitor ciprofloxacin in the treatment of *Bacillus anthracis* infections.

The nature of the DNA breaks generated during the strand transfer reaction led to a classification of DNA topoisomerases into two types: type I DNA topoisomerases (Topo I) introduce single-strand breaks in the sugar-phosphate backbone, whereas type II DNA topoisomerases (Topo II) introduce double-strand breaks. This nomenclature is complicated by the number added for some individual enzymes, according to the timing of their discovery. However, by chance, all type I DNA topoisomerases have an odd number (I, III, V) whilst type II DNA topoisomerases have an even number (II, IV, VI) (Table 1). Finally, the two DNA topoisomerases with supercoiling activities are usually named according to their unique characteristic, gyrase (a Topo II that produces negative supercoiling) and reverse gyrase (a Topo I that produces positive supercoiling).

In addition to the nature of the DNA break, Topo I and Topo II differ in their ATP dependence and quaternary structure. All Topo II are ATP dependent, whereas Topo I are ATP independent, with the exception of the reverse gyrase, for which an ATP requirement is due to the combined action of its associated helicase domain. Finally, whereas Topo I are monomeric, Topo II are multimeric with a dyad

**Table 1.** Classification of DNA topoisomerases

| Family | Linked | Enzyme | Domain | Activity | Representative species |
|---|---|---|---|---|---|
| | | topoisomerase I | B | rel − | *Escherichia coli* |
| | | topoisomerase II | B | rel − | *Escherichia coli* |
| IA | 5′ | topoisomerase III | E | rel − | *Scaccharomyces cerevisiae* |
| | | topoisomerase III | A | rel − | *Desulfurococcus* |
| | | topoisomerase III (α, β) (1) | E | rel − | *Homo sapiens* |
| | | reverse gyrase | A(B) | sup + | *sulfolobus acidocaldarius* |
| IB | | topoisomerase IB (2) | E | rel −+ | *Homo sapiens* |
| | 3′ | topoisomerase IB | B | rel −+ | *Deinococccus radiodurans* |
| IB(3) | | topoisomerase V | A(4) | rel −+ | *Methanopyrus kandleri* |
| IB | | topoisomerase I of eukaryal virus | V | rel −+ | *Vaccinia virus* |
| | | topoisomerase II | E | rel −+ | *Saccharomyces cerevisiae* |
| | | topoisomerase II (α, β) (1) | E | rel −+ | *Homo sapiens* |
| IIA | 5′ | topoisomerase II of eukaryal virus | V | rel −+ | *PBCV1* |
| | | topoisomerase IV | B | rel −+ | *Escherichia coli* |
| | | gyrase | B(A,E) | sup − | *Escherichia coli* |
| | | topoisomerase II of phages | V | rel −+ | *phage T4* |
| IIB | 5′ | topoisomerase VI | AE(5) | rel −+ | *Sulfolobus shibatae* |

The numbers, from I to VI, correspond to the chronology of the DNA topoisomerases discovery in *E. coli* (I, II, III, IV) or Archaea (V, VI). All enzymes with even number are of type II, and all enzymes with uneven number are of type I. The presence of the corresponding topoisomerase in a Domain is indicated by A for Archaea, B for Bacteria, E for Eukarya and V for Virus. Letters in brackets indicates that the corresponding Topo II has been transferred in this domain from another one. Activities: rel, relaxation; sup, supercoiling; −, negative superturns; +, positive superturns. Representative species are those from which the corresponding topoisomerase have been biochemically characterised in more details. (1) In mammals, two isoforms have been described for the topoisomerase II and III. (2) The "B" suffix is used to distinguish these enzymes from the type IA enzymes. (3) The sequences similarities with the others type IB DNA topoisomerases is weak, the topoisomerase V could be a member of a third DNA topoisomerase I family.[61] (4) Topo V has been only detected in the archaeon *Methanopyrus kandleri*. (5) Presently, complete eukaryal Topo VI are only known in plants.

symmetry, both DNA strands at the double-strand break being transiently linked to two identical subunits of the enzyme (e.g. the two GyrA subunits in the case of DNA gyrase, which is an heterotetramer of two GyrA and two GyrB subunits).

With the growing number of DNA topoisomerases discovered in the three domains of life (*Archaea, Bacteria* and *Eukarya*), a subdivision in A and B families within each type has finally been established, based on sequence similarities (Topo IA and IB, Topo IIA and IIB) (Table 1). Topo IA and Topo IB are non-homologous, sharing neither sequence nor structural similarities. Furthermore, they exhibit critical differences in their mechanism of action. In particular, Topo IA are transiently linked to the 5′ end of the single-strand DNA break, whereas Topo IB are transiently linked to the 3′ end. On the contrary Topo IIA and Topo IIB are somehow related, since they share a homologous ATP-binding domain (see below), and they are both transiently linked to the 5′ end of the double-strand DNA breaks.

Prototypes of the Topo IA, IB and IIA families were all discovered and first characterised in the seventies. In contrast, the Topo IIB family was only recognised in 1997, following the sequencing of the genes encoding the two subunits of the Topo II from the archaeon *Sulfolobus shibatae*.[4] The discovery of this atypical Topo II (named DNA topoisomerase VI) has raised new questions about the origin and evolution of

this class of enzymes: how many times has Topo II activity originated? Which Topo II (if any) were present in the last common cellular ancestor (LUCA)? Do Topo II phylogenies help to resolve the topology of the universal tree of life? These questions are important, since the origin of Topo II is most likely strongly connected to the origin of double-stranded DNA genomes. Trying to answer some of them, we review here briefly the properties of the two Topo II families and we present a phylogenomic analysis of their distribution in the three domains of life.

### The two families of type II DNA topoisomerases, Topo IIA and Topo IIB

Topo II are essential in all living cells since they are the only enzymes that can decatenate or unlink intact DNA duplexes. Accordingly, they are absolutely required at the stage of chromosome segregation to untangle newly replicated double-stranded DNA molecules. In addition, Topo II are also required to remove positive superturns ahead of replication forks and moving transcription complexes in Archaea and Bacteria, since the Topo IA present in these two domains cannot relax positive superturns (in Eukarya, these tasks can be performed by Topo IB as well) (Table 1). In addition, DNA gyrase (a Topo IIA) is essential in Bacteria, since it is responsible for the negative supercoiling of the chromosomes.

All members of the type IIA family exhibit extensive sequence similarities despite quaternary structure differences (Fig. 1). The two known bacterial Topo IIA (gyrase and TopoIV) are heterotetramers composed of two subunits: the TopoIV subunits, ParE and ParC, being homologues to the gyrase subunits GyrB and GyrA, respectively. These subunits are the targets of very important antibiotics (coumarins for GyrB/ParE and quinolones for GyrA/ParC).

Eukaryal Topo IIA are distinct from their prokaryotic homologues in term of structure. These enzymes are homodimers; their N-terminal half ends are homologous to GyrB/ParE subunits, whereas their central parts are homologous to GyrA/ParC subunits. Despite their structural and sequence similarities with their bacterial counterparts, eukaryal Topo IIA are relatively insensitive to coumarins and resistant to most quinolones. However, they are the targets of several important families of antitumoral drugs, such as epipodophylotoxines or amsacrides.

The C-terminal portions of eukaryal Topo IIA, which are already highly divergent between enzymes from different Eukarya, have no detectable similarities with the C-terminal regions of prokaryotic Topo IIA. This region is the target of regulatory processes in Eukarya (e.g. phosphorylation).[5]

TopoVI is presently the only member of the type IIB DNA topoisomerase family. This enzyme is an heterotetramer, composed of two subunits A and B, without significant simi-larities with enzymes of the Topo IIA family. Interestingly, the A subunit of TopoVI is homologous to the eukaryotic Spo11 protein.[4] This relationship led to an important break-through in the understanding of the molecular basis of meiosis with the identification of Spo11. This enzyme is responsible for the formation of the chromosomal double-strand breaks that initiate meiotic recombination in the yeast *Saccharomyces cerevisiae*.[4,6] Experimental evidence also implicates a role for Spo11 during meiosis in *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, and *Arabidopsis thaliana* (for a review on Spo11 see Ref. 7).

Both Topo IIA and Topo IIB share three short amino-acid motives located in the N-terminal region of their B subunits (Fig. 1). However, these motives are not Topo II specific, since they are present at the N-terminal domain of heat-shock proteins of the Hsp90 family[8] and repair proteins of the MutL family,[9] or at the C-terminal domain of histidine kinases.[10] As initially shown by site-directed mutagenesis on *E. coli* gyrase, they contain the conserved amino acids involved in ATP binding and hydrolysis.[11] These motives are now recognised as the signature of a new type of ATP module referred to as the "Bergerat fold".[12] Along with type IIA topoisomerases, all these enzymes are now grouped in the GHKL protein family.[12]
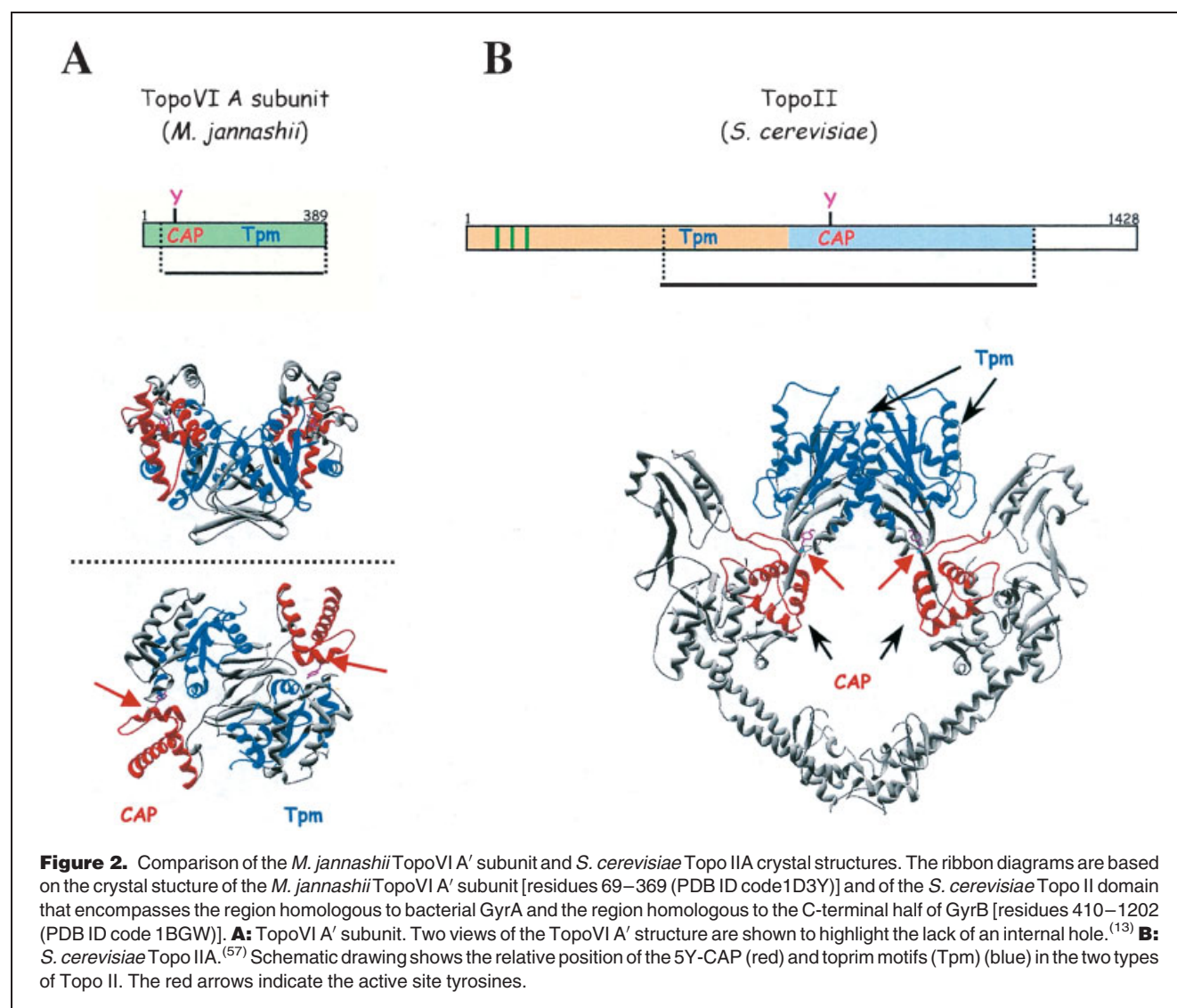
Whereas the B subunits of Topo IIA and Topo IIB share at least a common ATP module, their A subunits are completely
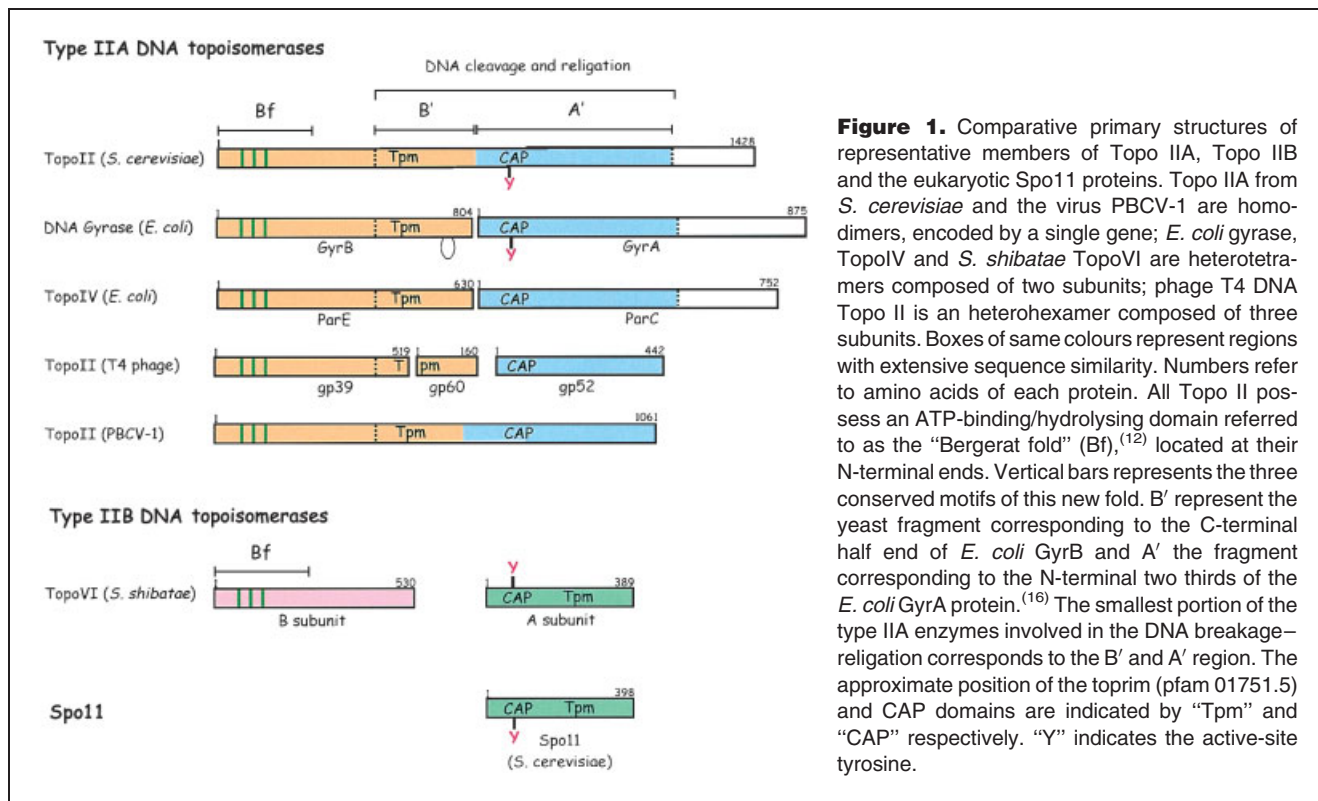


**Figure 1.** Comparative primary structures of representative members of Topo IIA, Topo IIB and the eukaryotic Spo11 proteins. Topo IIA from *S. cerevisiae* and the virus PBCV-1 are homo-dimers, encoded by a single gene; *E. coli* gyrase, TopoIV and *S. shibatae* TopoVI are heterotetra-mers composed of two subunits; phage T4 DNA Topo II is an heterohexamer composed of three subunits. Boxes of same colours represent regions with extensive sequence similarity. Numbers refer to amino acids of each protein. All Topo II possess an ATP-binding/hydrolysing domain referred to as the "Bergerat fold" (Bf),[12] located at their N-terminal ends. Vertical bars represents the three conserved motifs of this new fold. B′ represent the yeast fragment corresponding to the C-terminal half end of *E. coli* GyrB and A′ the fragment corresponding to the N-terminal two thirds of the *E. coli* GyrA protein.[16] The smallest portion of the type IIA enzymes involved in the DNA breakage–religation corresponds to the B′ and A′ region. The approximate position of the toprim (pfam 01751.5) and CAP domains are indicated by "Tpm" and "CAP" respectively. "Y" indicates the active-site tyrosine.

unrelated. This was first deduced from the absence of sequence similarities between them,[4] and nicely confirmed by structural analysis of a large proteolysis fragment of the archaeon *Methanococcus jannaschii* TopoVI A subunit (Fig. 2).[13] A CAP-like DNA-binding domain, which contains the active site tyrosine, is present on the proteolysis fragment termed A′ of type IIA enzyme and on TopoVI A subunit. However, this common feature probably does not correspond to an homologous trait, since the tyrosine is located in a β-β turn in Topo IIA,[14] but in an α-helix in *M. jannaschii* TopoVI.[13] Topo IIA and IIB also contain a subclass of the Rossman fold known as the toprim domain present in type IA DNA topoisomerases and other nucleic acid modifying enzymes.[15] This fold is located in the B subunit of Topo IIA, but in the A subunit of TopoVI (Figs. 1, 2).

Several important differences in the mechanisms of action between the two families have recently been found.[17] Although both Topo IIA and Topo IIB are transiently linked to the 5′ end of the double-strand DNA breaks, the structures of these breaks differ between both families: Topo IIA produces breaks with four base pairs overhangs,[16] whereas *S. shibatae* TopoVI generates breaks with two base pairs overhangs.[17] Secondly, DNA cleavage by TopoVI occurs only in the presence of ATP, whereas DNA cleavage by Topo IIA is ATP independent.[17] The presence of the CAP-like domain and the toprim fold are necessary and sufficient for the cleavage reaction by Topo IIA, since expression of a fragment of *S. cerevisiae* Topo IIA deleted of its ATP-binding domain (the Bergerat fold) is still able to catalyse the DNA cleavage reaction.[15] This is not the case for TopoVI, since no DNA



**Figure 2.** Comparison of the *M. jannaschii* TopoVI A′ subunit and *S. cerevisiae* Topo IIA crystal structures. The ribbon diagrams are based on the crystal stucture of the *M. jannaschii* TopoVI A′ subunit [residues 69–369 (PDB ID code1D3Y)] and of the *S. cerevisiae* Topo II domain that encompasses the region homologous to bacterial GyrA and the region homologous to the C-terminal half of GyrB [residues 410–1202 (PDB ID code 1BGW)]. **A:** TopoVI A′ subunit. Two views of the TopoVI A′ structure are shown to highlight the lack of an internal hole.[13] **B:** *S. cerevisiae* Topo IIA.[57] Schematic drawing shows the relative position of the 5Y-CAP (red) and toprim motifs (Tpm) (blue) in the two types of Topo II. The red arrows indicate the active site tyrosines.

cleavage has been observed with the A subunit of *S. shibatae* TopoVI alone,[18] despite the fact that both the CAP-like domain and the toprim folds are present on the TopoVI A subunit (Figs. 1, 2).

All these data suggest that Topo IIA and Topo IIB enzymes have been created independently by the assembly of several modules, some of them being common (the Bergerat fold, the CAP-like domain and the toprim domain), but other being unrelated. Thus, the Topo II activity has been most likely invented at least twice in the course of evolution, a situation reminiscent of many other enzymes involved in DNA metabolism (e.g. DNA polymerases, primases, DNA helicases, thymidylate synthases and so on).[19]

### Evolution of the two bacterial Topo IIA

All Topo II from Bacteria belong to the A family. This is quite significant, considering the number and variety of bacterial genomes that have now been completely or partly sequenced. Two different bacterial Topo IIA have been distinguished on the basis of biochemical studies in *E. coli*: gyrase and TopoIV. Only gyrase can introduce negative supercoiling into DNA. This unique feature is due to its ability to wrap a DNA segment around itself into a positive superturn. This wrapping induces the formation of a compensatory negative superturn into the topologically closed enzyme-free DNA. The positive superturn formed around gyrase is then converted into a negatively supercoiled one by the strand-transfer reaction of the enzyme, producing two negative superturns per cycle of topoisomerisation.[20] Wrapping of DNA around gyrase is critically dependent on the C-terminal region of its A subunit, since removal of this region converts DNA gyrase into a conventional Topo II.[21]

Beside their critical differences in term of gyrase activity, *E. coli* gyrase and TopoIV exhibit in vitro all canonical topoisomerase activities, i.e: catenation/decatenation, knotting/unknotting. Both enzymes are involved in removing topological stresses associated with replication and transcription by relaxing positive superturns in front of the replication forks and transcription complexes.[22] They also cooperate in decatenation of newly replicated DNA, the gyrase providing the negative supercoiling that stimulates the decatenation activity of TopoIV.[23,24] However, they have specialised roles *in vivo* in other aspects of DNA metabolism. In the case of *E. coli*, gyrase and TopoIV have antagonistic functions in maintaining the homeostasis of supercoiling in the cell. Whereas gyrase produces the negative supercoiling required for initiation of DNA replication and efficient initiation of transcription at supercoiling-dependent promoters, TopoIV cooperates with type I DNA topoisomerases to relax the excess of negative superturns introduced by gyrase.[25]

To better understand the relationships between gyrase and TopoIV, as well as their relationships with other Topo IIA, we have performed several phylogenetic analyses using either individual genes [(GyrA/ParC) and (GyrB/ParE) subunits], or concatenation of the GyrA/ParC and GyrB/ParE subunits. The right tree of Fig. 3 has been obtained with the fusion proteins, and includes viral and eukaryal Topo IIA, in order to root the bacterial part of the Topo II tree. Trees of similar topologies were obtained with the individual subunits.

Bacterial Topo IIA with demonstrated gyrase activity (marked with dollars in Fig. 3) are included into a monophyletic group (indicated by an arrow) emerging late in the bacterial Topo IIA tree (rooted in-between Bacteria and Eukarya), whereas all Topo IIA with a biochemically demonstrated TopoIV activity (marked with a hash in Fig. 3) are included into several groups of paraphyletic sequences that have a basal position. The monophyly of the gyrase cluster is not strongly supported and the assignment of a few sequences, either to the gyrase or the TopoIV cluster, is ambiguous (e.g. *Deinococcus*). It cannot be completely ruled out that TopoIV sequences are clustered artificially together, because their branches are slightly longer than those of gyrases. However, this could not be due to attraction by the long branches of the outgroup, since we also obtained a separation of prokaryotic Topo IIA into a single gyrase-like group and several TopoIV groups when we constructed an unrooted tree using only prokaryotic sequences, allowing to use 418 amino-acid positions instead of 294 (see supplementary materials at www-archbac.u-psud.fr/Projects/phylotopo/topo_debut.htm).

Most Topo II present in the ''gyrase group'' contain the conserved short amino-acid sequence, known as the ''GyrA box'', previously identified by Ward et al.[26] (see supplementary materials at www-archbac.u-psud.fr/Projects/phylotopo/topo_debut.htm). These authors suggested that this motif could be a signature for GyrA subunits, which is in agreement with the present data. In contrast to members of the TopoIV group, members of the gyrase group can be aligned in their C-terminal regions (not shown), in accordance with the importance of this region in the mechanism of gyration and with a previous analysis on a smaller number of sequences.[27] These results suggest that a simple phylogenetic analysis can generally determine whether bacterial Topo IIA is a gyrase or a TopoIV.

The early emergence of TopoIV lineages in the bacterial Topo IIA tree suggests that TopoIV is reminiscent of the ancestral bacterial Topo IIA, and that gyrase originated from TopoIV. In agreement with the antiquity of TopoIV, this enzyme is present in different unrelated bacterial phyla (*Proteobacteria*, *Firmicutes*, *Chlamidiae*, and *Spirochaetes*). The emergence of gyrase from TopoIV makes sense because the invention of gyrase activity was clearly a novelty that could have been selected in organisms already containing a decatenase (TopoIV). In contrast, the introduction and maintenance of a TopoIV in an organism already containing a gyrase with decatenase activity appears less likely. Interestingly, all Bacteria whose genomes have been completely
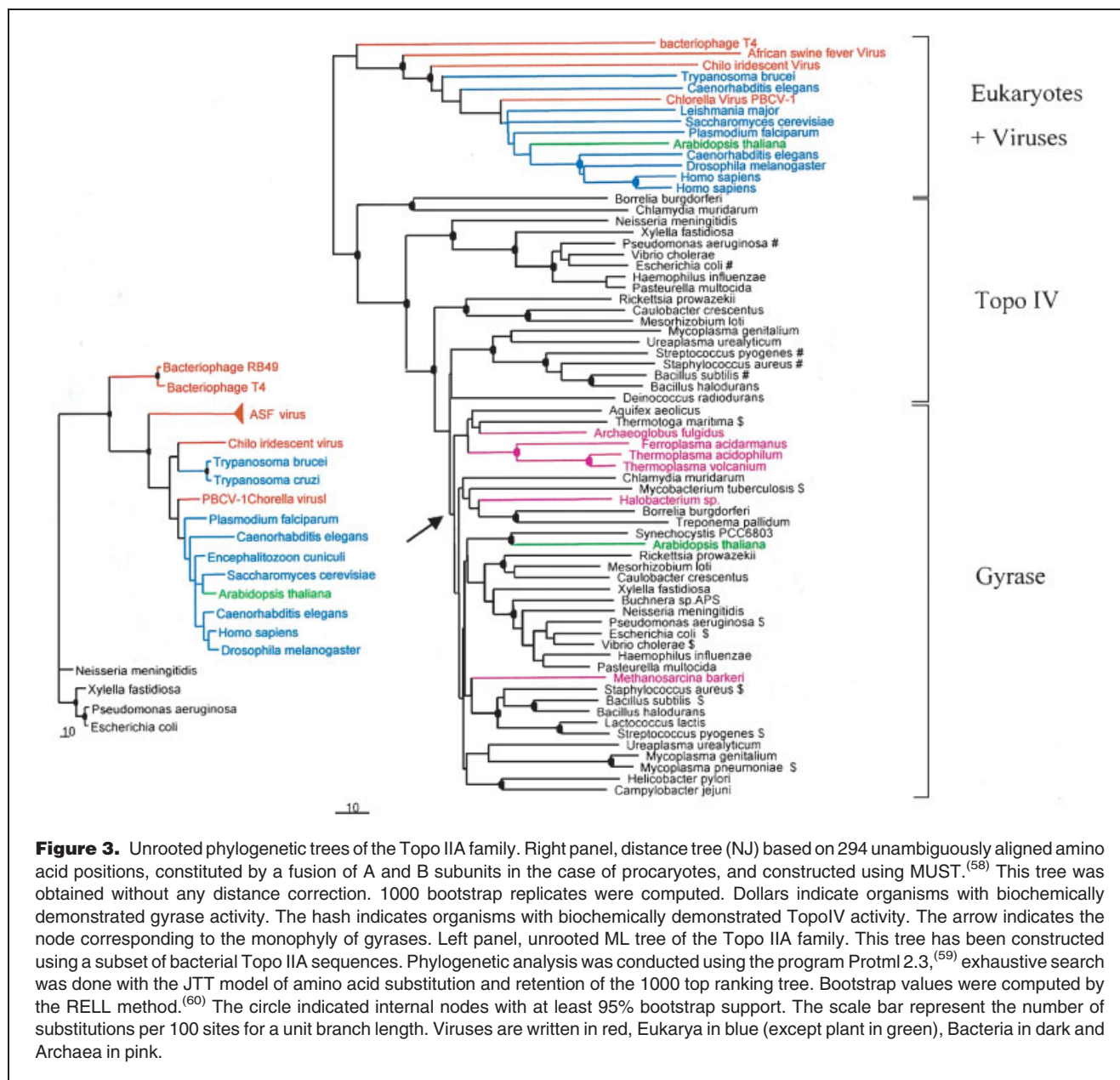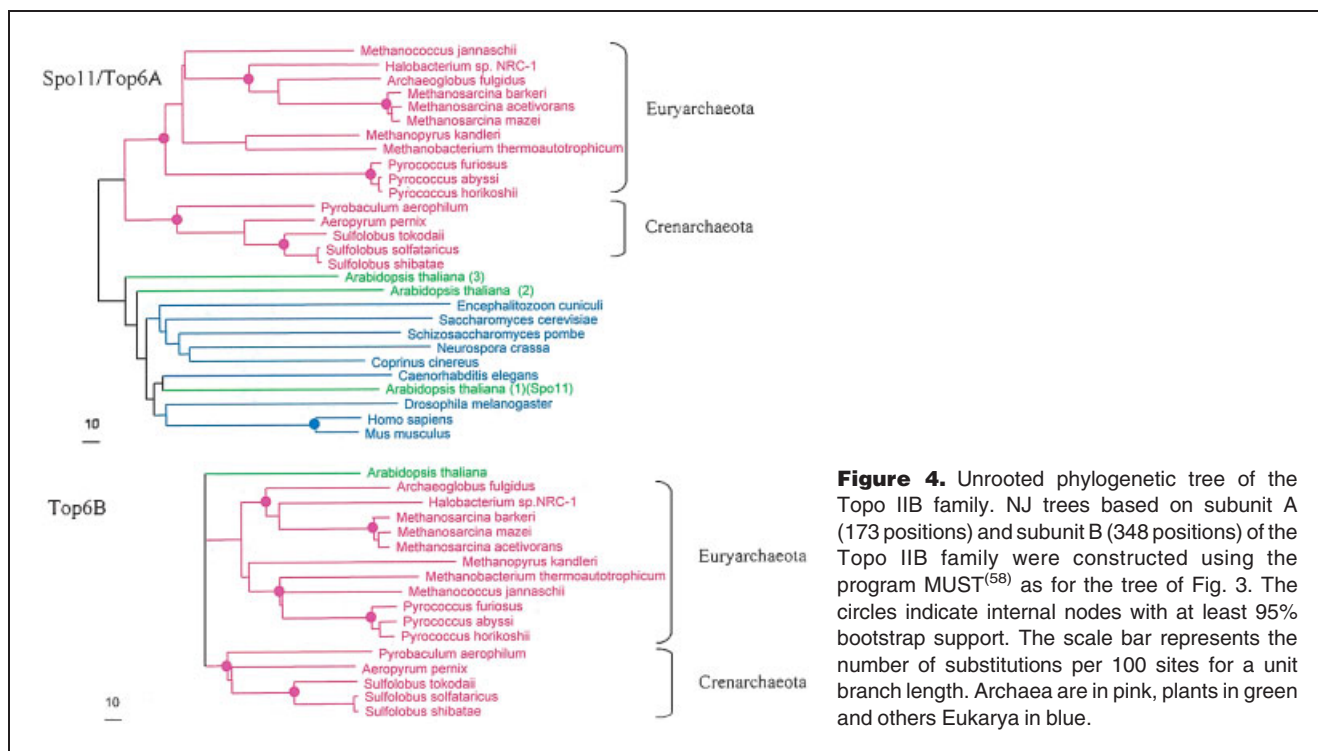
**Figure 3.** Unrooted phylogenetic trees of the Topo IIA family. Right panel, distance tree (NJ) based on 294 unambiguously aligned amino acid positions, constituted by a fusion of A and B subunits in the case of procaryotes, and constructed using MUST.[58] This tree was obtained without any distance correction. 1000 bootstrap replicates were computed. Dollars indicate organisms with biochemically demonstrated gyrase activity. The hash indicates organisms with biochemically demonstrated TopoIV activity. The arrow indicates the node corresponding to the monophyly of gyrases. Left panel, unrooted ML tree of the Topo IIA family. This tree has been constructed using a subset of bacterial Topo IIA sequences. Phylogenetic analysis was conducted using the program Protml 2.3,[59] exhaustive search was done with the JTT model of amino acid substitution and retention of the 1000 top ranking tree. Bootstrap values were computed by the RELL method.[60] The circle indicated internal nodes with at least 95% bootstrap support. The scale bar represent the number of substitutions per 100 sites for a unit branch length. Viruses are written in red, Eukarya in blue (except plant in green), Bacteria in dark and Archaea in pink.

sequenced so far contain a Topo IIA phylogenetically related to gyrase, whereas several of them lack sequences related to TopoIV. This suggests that both gyrase and TopoIV were present in the last bacterial ancestor, TopoIV having been lost later on independently in various lineages. The existence of many bacteria lacking TopoIV indicates that gyrase is a priori sufficient to solve all topological problems related to chromosome segregation, transcription and replication fork movements. The ubiquity of gyrase in the bacterial domain also indicates that gyrase activity is critical for Bacteria. One can imagine that gyrase became essential once mechanisms dealing with DNA in Bacteria became dependent on negative

supercoiling. Possibly, the organism that first acquired a gyrase was the ancestor of all present-day Bacteria.

## Both type II DNA topoisomerases families are present in Archaea

In contrast to the situation observed in Bacteria, both Topo IIA and Topo IIB are present in Archaea. Topo IIB is widespread, since its two genes have been detected in all Archaea whose genomes have been completely sequenced, with the only exception of *Thermoplasmatales*. This suggests that Topo IIB was already present in the last common ancestor of all Archaea. In agreement with this hypothesis, the phylogenies

**Figure 4.** Unrooted phylogenetic tree of the Topo IIB family. NJ trees based on subunit A (173 positions) and subunit B (348 positions) of the Topo IIB family were constructed using the program MUST[58] as for the tree of Fig. 3. The circles indicate internal nodes with at least 95% bootstrap support. The scale bar represents the number of substitutions per 100 sites for a unit branch length. Archaea are in pink, plants in green and others Eukarya in blue.

of both archaeal TopoVI A and B subunits (Fig. 4) resemble the archaeal phylogeny based on 16 S rRNA tree, with a clear distinction of the *Euryarchaeota* and *Crenarchaeota* phyla. In particular, the phylogeny of the A subunit is similar to those obtained using a concatenation of ribosomal proteins, with the early emergence of *Pyrococcus* in the euryarchaeal branch, and a clade grouping *Halobacteria*, *Methanosarcina* and *Archaeoglobus*.[28]

Unlike Topo IIB, Topo IIA have a restricted distribution in Archaea, being only present in a subset of *Euryarchaeota* (the three *Thermoplasmatales*, *Halobacterium* NRC1, *Archaeoglobus fulgidus* and *Methanosarcina mazei*). This can be explained either by multiple loss of Topo IIA during archaeal evolution, or by Lateral Gene Transfer (LGT). The branching of all archaeal Topo IIA among bacterial gyrases in the Topo IIA tree (Fig. 3) supports the latter hypothesis (indicating a transfer from Bacteria to Archaea) and further suggests that archaeal Topo IIA have the gyrase activity (all archaeal Topo IIA also contain the GyrA box previously mentioned). Although archaeal Topo IIA have not yet been characterised biochemically, they are most likely gyrase, since their presence correlates with the negative supercoiling of resident plasmids in both *Haloarchaea* and *Archaeoglobus profondus*[29,30] whereas plasmids isolated from Archaea lacking Topo IIA genes are either relaxed or positively supercoiled.[31] The gyrase of *Haloarchaea* is likely essential for the maintenance of in vivo negative supercoiling, since haloarchaeal plasmids become positively supercoiled after

treatment with novobiocin, a classical gyrase inhibitor.[29] Evidence that halobacterial gyrase is the target of novobiocin has been obtained from genetic analysis, as a novobiocin-resistant mutant of *Haloferax alicantei* was mapped in its Topo IIA *gyrB*-like gene.[32]

The non-orthologous displacement of TopoVI by bacterial-like gyrases in the three *Thermoplasmatales* indicates that gyrase can perform in these euryarchaeon all functions normally done by TopoVI in other Archaea (relaxation of positive supercoiling, decatenation and unknotting) as in Bacteria lacking TopoIV. On the contrary, the presence of both TopoVI and gyrase in *A. fulgidus*, *Halobacterium* NRC1 and the *Methanosarcinales* suggests that these two enzymes are specialised to perform different subsets of functions in those Archaea, as in Bacteria containing both gyrase and TopoIV. Interestingly, the four genes encoding the two subunits of TopoVI and the two subunits of gyrase are located next to each other in the genome of *Halobacterium* NRC1, suggesting a common regulation.

Archaeal gyrases are present in mesophiles (*Methanosarcinales* and *Halophiles*), three moderate thermophiles (*Thermoplasmatales*) and a hyperthermophile (*A. fulgidus*), indicating the absence of correlation between the optimal growth temperature and gyrase activity (Fig. 3). This is strikingly different from the situation observed in the case of reverse gyrase, which is only present in hyperthermophiles.[33,34] Topological analysis of resident plasmids in *T. maritima* and *Archaeoglobus profondus*, which contain

both gyrase and reverse gyrase, have shown that gyrase activity predominates above reverse gyrase in defining the global supercoiled state of plasmids, as these plasmids turned out to be negatively supercoiled.[27,30]

## Eukaryal Topo IIA and Spo11

Eukaryal Topo IIA are mainly involved in chromosome segregation and condensation via their decatenation activity.[35] They are probably also involved in the relaxation of positive superturns produced by the progression of transcription and replication complexes along the DNA. The latter tasks can also be a priori performed by the eukaryal Topo IB, which, unlike Topo IA, can relax positive superturns (Table 1). In the case of *Schizosaccharomyces pombe*, Topo IB and Topo IIA can indeed complement each other.[36] However, in mammals, Topo II$\alpha$ is probably directly involved in transcription, since its activity is essential for in vitro transcription of a template assembled into chromatin by RNA polymerase II holoenzyme.[37]

All Eukarya contain at least one member of the Topo IIA family that form a monophyletic cluster well separated from prokaryotic Topo IIA in the tree of the complete Topo IIA family (Figure 3). This strongly suggests that this eukaryal subtype of Topo IIA was already present in the last common ancestor of all Eukarya. The phylogenetic tree of these eukaryal Topo IIA is highly asymmetric, suggesting that it is prone to reconstruction artefacts (Fig. 3).[38] Indeed, *A. thaliana* (a plant) branches in-between animals and yeast, in contrast to the sisterhood of animal and fungi in the 18 S rRNA tree. The three protists lineages emerge with long branches at the base of the tree, together with a Topo IIA from *Caenorhabditis elegans* and several Topo IIA encoded by eukaryal viruses (see below). The presence of two divergent Topo IIA in *C. elegans* can be explained either by an early gene duplication of the Topo IIA gene in Eukarya, followed by multiple gene loss, or more likely by duplication of the Topo IIA gene in a *C. elegans* ancestor, followed by rapid evolution of one of the two copies.

A surprising outcome of our BLAST search for all putative Topo IIA sequences was the identification of two Topo IIA in *Arabidopsis thaliana*: one groups with eukaryal Topo IIA, as expected, but the other groups with bacterial gyrases. This novel *A. thaliana* Topo IIA, which is the first gyrase-like protein detected in Eukarya, seems to be widespread in plants, since partial sequences can be detected in several ongoing genome projects (unpublished observation). The enzyme most likely originated from the bacterial ancestor of chloroplasts, being sister lineage of cyanobacterial gyrase with a bootstrap value of 100% (Fig. 3).

In addition to Topo IIA, all completely sequenced eukaryal genomes contain at least one homologue of the archaeal TopoVI A subunit, the Spo11 protein involved in meiotic recombination. Since no homologs of the B subunit of archaeal

TopoVI are present in completely sequenced genomes from fungi and animals, Spo11 protein either evolved the ability to introduce double-strand break alone in meiotic recombination, or recruited another protein to replace the B subunit in Eukarya. Archaeal TopoVI A subunits and eukaryal Spo11 are clearly separated in our phylogenetic analysis (Fig. 4). The Spo11 phylogeny fits the eukaryotic 18S rRNA tree, with yeast and animals clustering together, indicating that these proteins are most likely orthologues and that Spo11 was present in the last common eukaryal ancestor.

Whereas yeast and animals encode only one Spo11 protein, the genome of *A. thaliana* encodes three Spo11 paralogues[39,40] (Fig. 4). Genetic analyses indicate that only one of the three Spo11 (AtSpo11-1) is involved in meiosis.[39] This protein branches with other eukaryal Spo11 in our phylogenetic tree (Fig. 4) while the others, AtSpo11-2 and AtSpo11-3, emerged at the base of the eukaryal tree. Recently, the sequencing of the *A. thaliana* genome has led, for the first time, to the detection of an eukaryal homologue of the archaeal TopoVI B subunit.[40] Two-hybrid studies have shown that the *A. thaliana* TopoVIB-like protein interacts in vivo with AtSpo11-2 and AtSpo11-3 but not with AtSpo11-1.[40] Genetic and physiological analyses of *A. thaliana* mutants located in *AtSpo11-3* and *AtTop6B* genes revealed identical phenotype affecting the expression of many genes.[41−43] The products of these two genes have also being recently directly implicated in endoreduplication, a chromosome amplification mechanism that controls cell and plant size via polyploïdy.[42,43] These data strongly suggest that *Spo11-3* encodes for the subunit A of a *bona fide* TopoVI in plants.

## Viral Topo II

Several viruses with large DNA genomes encode for their own Topo II (presently all from the A family with T4 Topo II the only one that has been extensively studied).[44] This enzyme, essential for the initiation of phage DNA replication,[45] is an heterohexamer of three subunits (Fig. 1). More recently, an heterodimeric type II enzyme of the Chlorella virus PBCV-1 has also been analysed in some details.[46] All known viral type II enzymes lack a C-terminal domain present in cellular type IIA enzymes, indicating that this region is dispensable for the transport of a DNA duplex through another.

Viral Topo IIA does not branch with the Topo IIA of their hosts in the Topo IIA phylogenetic tree (Fig. 3). Bacteriophage T4 Topo IIA branches in-between bacterial and eukaryal Topo IIA, whereas Topo IIA from African Swine Fever Virus (ASFV) and Chilio iridescent Virus (CIV) branch in-between T4 and eukaryal Topo IIA. The third Topo IIA encoded by an eukaryal virus (PBCV-1) branches after two Topo IIA protists, but before the crown of the eukaryal Topo IIA tree that includes animals, plants and fungi.

Considering the asymmetric shape of the viral and eukaryal Topo IIA tree, the clustering of most viral sequences together

could be due to differences in evolutionary rates between viral and cellular sequences. We have thus performed a Maximum Likelihood (ML) analysis of the Topo IIA phylogeny, since this method is less sensitive to differences in evolutionary rates.[47] In the ML tree, T4 and ASFV sequences are still in-between bacterial and eukaryal ones, whereas CIV now branch with *Trypanosoma* (Fig. 3, left panel). This result suggests an ancient origin for some viral Topo II (predating the diversification of the bacterial and eukaryal domains). However, the position of viral sequences in the Topo IIA tree cannot be considered as definitive, since it is known that, in some cases, ML analysis are insufficient to correct artefacts due to long branch attraction.[48]
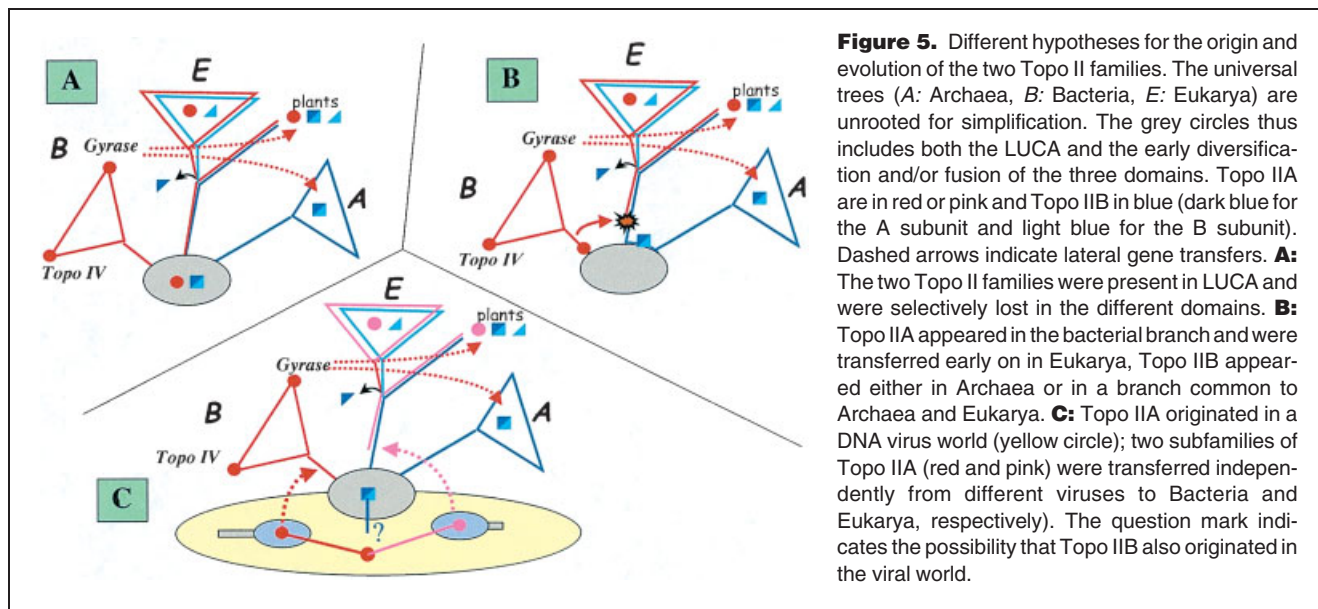
### Evolutionary perspective

The two families of Topo II exhibit strikingly different distributions among the three domains of life. Topo IIA is widespread, being ubiquitous in Bacteria and Eukarya, and present in some *Euryarchaea* and viruses. Conversely, Topo IIB are mostly present in Archaea and plants, but completely absent in Bacteria. Considering that archaeal Topo IIA likely originated from *Bacteria* by several independent lateral gene transfers (see Fig. 3), there is a clear-cut distribution of both families in each prokaryotic domain: Topo IIA and Topo IIB being associated with Bacteria and Archaea, respectively. This suggests that the last common ancestor of Archaea contained only Topo IIB, whilst the last common ancestor of Bacteria contained only Topo IIA.

The situation in Eukarya is more puzzling, as outlined by the case of plants where both Topo II families coexist. The last common ancestor of Eukarya certainly contained a Topo IIA and at least the A subunit of Topo IIB (Spo11). It is not possible to determine if the B subunit was initially also present in

Eukarya and later on lost in animals and fungi, or if the B subunit in plants has been acquired from Archaea by LGT. Further sequencing of complete genomes from various protists will hopefully help to answer this question.

Eukaryal genomes contain many "bacterial genes" that have been acquired via mitochondria or chloroplasts. This should be the case for the plant gyrase-like Topo IIA detected in this work (see Fig. 3). This excludes a priori the same origin for the classical eukaryal Topo IIA. Indeed, whilst the gyrase-like Topo IIA in plants cannot be distinguished at the sequence level from the bacterial enzymes, the classical eukaryal Topo IIA are highly divergent from all bacterial Topo IIA, both TopoIV and gyrases. Such divergence cannot be explained a priori by a change in function or in cellular context, since bacterial gyrase and TopoIV are much less divergent, despite radical change in function, whereas plants, bacterial and archaeal gyrases appear to be very similar, in spite of different cellular contexts. As a consequence, either Topo IIA was originally present in LUCA or it was transferred very early on from Bacteria to Eukarya (i.e. before any eukaryal diversification).

As previously noticed, the existence of two non-homologous families of proteins performing the same function in DNA metabolism is not specific for Topo II.[19] In particular, a striking observation from comparative genomic is that most DNA replication proteins are not homologous in Bacteria and Archaea.[49] This is surprising considering that all cellular DNA replication systems are very similar in term of general mechanism and enzymatic functions involved. Three hypotheses, which are relevant to the origin of Topo II, have been proposed to explain this paradox.[49–51] The first possibility (Fig. 5A) is that two non-homologous replication machineries coexisted in an already elaborated LUCA and that only one was later on conserved in Archaea and Bacteria. In that case,



**Figure 5.** Different hypotheses for the origin and evolution of the two Topo II families. The universal trees (*A:* Archaea, *B:* Bacteria, *E:* Eukarya) are unrooted for simplification. The grey circles thus includes both the LUCA and the early diversification and/or fusion of the three domains. Topo IIA are in red or pink and Topo IIB in blue (dark blue for the A subunit and light blue for the B subunit). Dashed arrows indicate lateral gene transfers. **A:** The two Topo II families were present in LUCA and were selectively lost in the different domains. **B:** Topo IIA appeared in the bacterial branch and were transferred early on in Eukarya, Topo IIB appeared either in Archaea or in a branch common to Archaea and Eukarya. **C:** Topo IIA originated in a DNA virus world (yellow circle); two subfamilies of Topo IIA (red and pink) were transferred independently from different viruses to Bacteria and Eukarya, respectively). The question mark indicates the possibility that Topo IIB also originated in the viral world.

both Topo IIA and Topo IIB were present in LUCA. The second hypothesis, illustrated in Fig. 5B, is that DNA replication was invented twice independently in Archaea and Bacteria. In that case, Topo IIA and Topo IIB also appeared independently in Bacteria and Archaea, respectively. This scenario is in agreement with the idea of a LUCA with an RNA genome[50] that did not required Topo II activity. Finally, it was suggested that some present-day cellular DNA replication proteins are of viral origin in Bacteria and Eukarya.[51−53] The latter scenario could explain why viral Topo IIA are clustered at the root of the Topo IIA tree of Fig. 3 if Bacteria and Eukarya recruited different versions of Topo IIA that had previously diverged in the viral world. The viral scenario is compatible with a LUCA containing one Topo II (Topo IIB) or no Topo II at all. In an extreme view, both Topo II could be of viral origin (question mark in Fig. 5C), in agreement with the recent hypothesis that both DNA and DNA replication proteins originated in the viral world.[19] We have not rooted the trees of Fig. 5 since phylogenetic relationships between the three domains of life are still highly debated (for recent reviews see Refs. 54−56). The distribution of Topo IIA and IIB can be indeed reconciled with competing proposals for the topology of the universal tree of life, assuming few ad hoc hypotheses of gene loss or transfer.

## Conclusions

The structural and phylogenomic analysis of both Topo II families strongly suggests that the Topo II activity appeared twice independently in early life evolution but did not allow a choice to be made between competing scenarios for the nature of LUCA, or for the early relationships between cells and viruses. However, one can hope that the discovery and analysis of new Topo II sequences in the viral and cellular worlds will finally permit us to determine which of these scenarios is the most likely. In the meantime, our analysis clearly emphasises the importance that Topo II could have played in early life evolution. For example, the distribution of bacterial Topo IIA subfamilies strongly suggests that the invention of DNA gyrase was critical in the formation of the bacterial domain. Considering the global effect of gyrase activity at the genome level, it is surprising that this enzyme has been successfully transferred several times in totally new cellular context (from Bacteria to Archaea, including some hyperthermophilic Archaea, and from Bacteria to plants). The analysis of DNA topology in organisms containing these ''foreign'' gyrases should be especially useful in understanding how LGT can affect the cellular machinery. In addition, considering the similar activities of TopoIV and TopoVI in vitro, and the extent of gene transfers that have occurred between the two prokaryotic domains, it is surprising that TopoIV has apparently never been transferred from Bacteria to Archaea whereas TopoVI has never been transferred from Archaea to Bacteria! This suggests that only transfers conferring novel selectable phenotypes can be successful. Alternatively,

TopoIV and/or TopoVI could interact in their respective domain with protein partners that are domain specific (possibly in connection with their roles in DNA replication and/or chromosome segregation), preventing their correct function in a foreign domain. Finally, the origin and evolution of Topo II appeared to be also linked to the origin of meiosis in Eukarya. The origin of sexuality is thus clearly deeply entrenched in our archaeal companions.

## Note added in proof

The structure of the B subunit of ths S. shibatae Topo VI has now been resolved (Corbett and Berger, EMBO J. 22, 151–163, 2003). The B subunit of Topo IIB and Topo IIA turned out to be structurally homologous, including their C-terminal regions, indicating homolgy. Thus, either Topo IIA and Topo IIB originated independently (as suggested in our review) by the recruitment of different A subunits to complement the same B subunit, or the A subunit of the ancestral family (Topo IIA or Topo IIB) was replaced by another non-homologous subunit to create a new family.

## References

1. Wang JC. DNA topoisomerases. Annu Rev Biochem 1996;65:635–692.
2. Champoux JJ. DNA topoisomerases:structure, function, and mechanism. Annu Rev Biochem 2001;70:369–413.
3. Li TK, Liu LF. Tumor cell death induced by topoisomerase-targeting drugs. Annu Rev Pharmacol Toxicol 2001;41:53–77.
4. Bergerat A, De Massy B, Gadelle D, Varoutas PC, Nicolas A, Forterre P. An atypical type II DNA topoisomerase from Archaea with implication for meiotic recombination. Nature 1997;386:414–417.
5. Austin CA, Marsh KL. Eukaryotic DNA topoisomerase II beta. Bioessays 1998;20:215–226.
6. Keeney S, Giroux CN, Kleckner N. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell 1997;88:375–384.
7. Keeney S. Mechanism and control of meiotic recombination initiation. Cur. Topics in Dev Biol 2001;52:1–55.
8. Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH. Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. Cell 1997;90:65–75.
9. Ban C, Yang W. Crystal structure and ATPase activity of MutL:implications for DNA repair and mutagenesis. Cell 1998;95:541–452.
10. Bilwes AM, Alex LA, Crane BR, Simon MI. Structure of CheA, a signal-transducing histidine kinase. Cell 1999;96:131–141.
11. Jackson AP, Maxwell A. Identifying the catalytic residue of the ATPase reaction of DNA gyrase. Proc Natl Acad Sci USA 1993;90:11232–11236.
12. Dutta R, Inouye M. GHKL, an emergent ATPase/kinase superfamily. Trends Biochem Sci 2000;25:24–28.
13. Nichols MD, DeAngelis K, Keck JL, Berger JM. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. EMBO J 1999;18:6177–6188.
14. Berger JM, Gamblin SJ, Harrison SC, Wang JC. Structure and mechanism of DNA topoisomerase II. Nature 1996;379:225–232.
15. Aravind L, Leipe DD, Koonin EV. Toprim:a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, old family nucleases and RecR proteins. Nucleic Acids Res 1998;26:4205–4213.
16. Wang JC. Moving one DNA double helix through another by a type II DNA topoisomerase:the story of a simple molecular machine. Q Rev Biophys 1998;31:107–144.
17. Buhler C, Lebbink J, Bocs C, Ladenstein R, Forterre P. DNA Topoisomerase VI Generates ATP Dependant Double-Strand Breaks with Two Nucleotide Overhangs. J Biol Chem 2001;276:37215–37222.
18. Buhler C, Gadelle D, Forterre P, Wang JC, Bergerat A. Reconstitution of DNA topoisomerase VI of the thermophilic archaeon *Sulfolobus shibatae*

from subunits separately overexpressed in *Escherichia coli*. Nucl acids Res 1998;26:5157–5162.

19. Forterre P. The origin of DNA and DNA genomes. Current Opin in Microbiol 2002;5:525–532.

20. Brown PO, Cozzarelli NR. A sign inversion mechanism for enzymatic supercoiling of DNA. Science 1979;206:1081–1083.

21. Kampranis SC, Maxwell A. Conversion of DNA gyrase into a conventional type II topoisomerase. Proc Natl Acad Sci USA 1996;93:14416–14421.

22. Khodursky AB, Peter BJ, Schmid MB, DeRisi J, Botstein D, Brown PO, Cozzarelli NR. Analysis of topoisomerase function in bacterial replication fork movement:use of DNA microarrays. Proc Natl Acad Sci USA 2000;97:9419–9424.

23. Zechiedrich EL, Cozzarelli NR. Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in *Escherichia coli*. Genes Dev 1995;9:2859–2869.

24. Zechiedrich EL, Khodursky AB, Cozzarelli NR. Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in *Escherichia coli*. Genes Dev 1997;11:2580–2592.

25. Zechiedrich EL, Khodursky AB, Bachellier S, Schneider R, Chen D, Lilley DM, Cozzarelli NR. Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. J Biol Chem 2000;275:8103–8113.

26. Ward DV, Newton A. Cell cycle expression and transcriptional regulation of DNA topoisomerase IV genes in *caulobacter*. J Bacteriol 1997;181:3321–3329.

27. Guipaud O, Marguet E, Noll K, Bouthier de la Tour C, Forterre P. Both DNA gyrase and reverse gyrase are present in the hyperthermophilic bacterium *Thermotoga maritima*. Proc Natl Acad Sci USA 1997;94:10606–10611.

28. Forterre P, Brochier C, Philippe H. Evolution of the Archaea. Theor Pop Biol 2002;61:409–422.

29. Sioud M, Baldacci G, De Recondo AM, Forterre P. Novobiocin induces positive supercoiling of small plasmids from halophilic archaebacteria in vivo. Nucleic Acids Res 1988;16:1379–1386.

30. Lopez-Garcia P, Forterre P, Van der Oost J, Erauso G. Plasmid pGS5 from the hyperthermophilic archaeon *Archaeoglobus profundus* is negatively supercoiled. J Bacteriol 2000;182:4998–5000.

31. Charbonnier F, Forterre P. Comparison of plasmid DNA topology among mesophilic and thermophilic eubacteria and archaebacteria. J Bacteriol 1994;176:1251–1259.

32. Holmes ML, Dyall-Smith ML. Mutations in DNA gyrase result in novobiocin resistance in halophilic archaebacteria. J Bacteriol 1991;173:642–648.

33. Forterre P, Bouthier De La Tour C, Philippe H, Duguet M. Reverse gyrase from hyperthermophiles:probable transfer of a thermoadaptation trait from archaea to bacteria. Trends Genet 2000;16:152–154.

34. Forterre P. A hot story from comparative genomics:reverse gyrase is the only hyperthermophile-specific protein. Trends Genet 2002;18:236–237.

35. Nitiss JL. Investigating the biological functions of DNA topoisomerases in eukaryotic cells. Biochim Biophys Acta 1998;1400:63–81.

36. Uemura T, Yanagida M. Isolation of type I and II DNA topoisomerase mutants from fission yeast:single and double mutants show different phenotypes in cell growth and chromatin organization. EMBO J 1984;3:1737–1744.

37. Mondal N, Parvin JD. DNA topoisomerase II alpha is required for RNA polymerase II transcription on chromatin templates. Nature 2001;413:435–438.

38. Philippe H, Laurent J. How good are deep phylogenetic trees? Curr Opin Genet Dev 1998;8:616–623.

39. Grelon M, Vezon D, Gendrot G, Pelletier G. AtSPO11-1 is necessary for efficient meiotic recombination in plants. EMBO J 2001;20:589–600.

40. Hartung F, Puchta H. Molecular characterization of homologues of both subunits A (SPO11) and B of the archaebacterial topoisomerase 6 in plants. Gene 2001;271:81–86.

41. Yin Y, Cheong H, Friedrichsen D, Zhao Y, Hu J, Mora-Garcia S, Chory J. A crucial role for the putative Arabidopsis topoisomerase VI in plant growth and development. Proc Natl Acad Sci USA 2002;99:10191–10196.

42. Sugimoto-Shirasu K, Stacey NJ, Corsar J, Roberts K, McCann M. DNA topoisomerase VI is essential for endoreduplication in *Arabidopsis*. Current Biol 2002;12:1–20.

43. Hartung F, Angelis KJ, Meister A, Schubert I, Melzer M, Puchta H. An archaebacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation in plants. Current Biol 2002;12:1787–1791.

44. O'Reilly EK, Kreuzer KN. Unique type II topoisomerase mutant that is hypersensitive to a broad range of cleavage-inducing antitumor agents. Biochemistry 2002;41:7989–7997.

45. Stetler GL, King GJ, Huang WM. T4 DNA-delay proteins, required for specific DNA replication, form a complex that has ATP-dependent DNA topoisomerase activity. Proc Natl Acad Sci USA 1979;76:3737–3741.

46. Lavrukhin OV, Fortune JM, Wood TG, Burbank DE, Van Etten JL, Osheroff N, Lloyd RS. Topoisomerase II from Chlorella virus PBCV-1. Characterization of the smallest known type II topoisomerase. J Biol Chem 2000;275:6915–6921.

47. Felsenstein J. Evolutionary trees from DNA sequences:a maximum likelihood approach. J Mol Evol 1981;17:368–376.

48. Wiens JJ, Servedio MR. Phylogenetic analysis and intraspecific variation:performance of parsimony, likelihood, and distance methods. Syst Biol 1998;4:7228–47253.

49. Edgell DR, Doolittle WF. Archaea and the origin(s) of DNA replication proteins. Cell 1997;89:995–998.

50. Leipe DD, Aravind L, Koonin EV. Did DNA replication evolve twice independently? Nucleic Acids Res 1999;27:3389–3401.

51. Forterre P. Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. Mol Microbiol 1999;33:457–465.

52. Villarreal LP, DeFilippis A. Hypothesis for DNA viruses as the origin of eukaryotic replication proteins. J Virol 2000;74:7079–7084.

53. Takemura M. Poxviruses and the origin of the eukaryotic nucleus. J Mol Evol 2001;52:419–425.

54. Forterre P, Philippe H. Where is the root of the universal tree of life? Bioessays 1999;21:871–879.

55. Gribaldo S, Philippe H. Ancient phylogenetic relationships. Theor Popul Biol 2002;61:391–408.

56. Lopez-Garcia P, Moreira D. Metabolic symbiosis at the origin of eukaryotes. Trends Biochem Sci 1999;24:88–93.

57. Berger JM, Gamblin SJ, Harrison SC, Wang JC. Structure and mechanism of DNA topoisomerase II. Nature 1996;379:225–232.

58. Philippe H. ''MUST, a computer package of Management Utilities for Sequences and Trees.'' Nucleic Acids Res 1993;21:5264–5272.

59. Adachi J, Hasegawa M. Molphy version 2.3 :programs for molecular phylogenetics based on maximum likelihood. Comput Sci Monogr 1996;28:1–150.

60. Kishino H, Miyata T, Hasegawa M. Maximum likelihood inference for protein phylogeny, and the origin of chloroplasts. J Mol Evol 1990;31:151–160.

61. Belova GI, Prasad R, Kozyavkin SA, Lake JA, Wilson SH, Slesarev AI. A type IB topoisomerase with DNA repair activities. Proc Natl Acad Sci USA 2001;98:6015–6020.