**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

_____ _____

Eduardo A. Torre Date

**Genome-wide Analysis of the DNATopoisomerase II and its Relationship to the CP190 Insulator Protein**

By

**Eduardo A. Torre**
Master of Science

Graduate Division of Biological and Biomedical Science
Molecular Biology

_____
**Victor G. Corces, PhD**
Advisor

_____
**John Lucchesi, PhD**

Committee Member

_____
**Robert McCauley, PhD**
Committee Member

_____
**William G. Kelly, PhD**
Committee Member

Accepted

_____
**Lisa A. Tedesco, PhD**
Dean of the James T. Laney School of Graduate Studies

_____
Date

# Genome-wide Analysis of the DNA Topoisomerase II and its Relationship to the CP190 Insulator Protein

By

**Eduardo A. Torre**

Advisor: Victor G. Corces, PhD

An abstract of a thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Master of Science in Biology

2011

Abstract

Topoisomerase II (Topo II) is an enzyme known to reduce DNA super-coiling by introducing negative coils in the double helix. It is involved in processes such as chromosome segregation, condensation, homologous recombination, DNA replication, and DNA transcription,. Chromatin immunoprecipitation (ChIP) followed by microarray analysis (ChIP-chip) experiments were done to determine the genome-wide binding pattern of Topo II. This study shows that Topo II binds to genomic intervals suggesting that it may bind DNA in a structure- specific rather than in a sequence-specific manner. In addition, I show that Centrosomal Protein 190 (CP190), a protein essential for the proper function of most known *Drosophila* insulators, delimits the distribution of this enzyme and that RNA interference- (RNAi) mediated knock down of CP190 leads to a change in the distribution pattern of Topo II. These data suggest that insulators may play an essential role in fine-tuning the distribution of Topo II and cells may use the controlled recruitment of CP190 as a means of regulating Topo II.

# Genome-wide Analysis of the DNA Topoisomerase II and its Relationship to the CP190 Insulator Protein

By

**Eduardo A. Torre**

Advisor: Victor G. Corces, PhD

A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Master of Science in Biology

2011

## TABLE OF CONTENTS



## TABLE OF IMAGES

# Introduction

Proper functioning of the genome requires complex handling of DNA molecules. This entails opening and closing of the DNA double helix, packing and unpacking of the chromatin structure, joining and separating sister chromatids, and all the processes required for replication, recombination, transcription and the formation of higher-order chromatin organization. Without successful operation of these processes genomic information will not be faithfully maintained or expressed.

In the 1960's Weil and Vinograd noted that some of the processes involved in management of DNA information could lead to topological problems where the double-stranded nature of the DNA molecule could cause over-winding in the very same way a rope over-winds once it is separated at one side and held together at the other (Weil, 1963). This over-winding, or super-coiling, could present a problem for molecular machineries to access their active sites within the genome. A solution to how nature deals with this problem arose with the discovery of a family of proteins known as the DNA topoisomerases. The first of them, known as Topoisomerase I (Topo I) was discovered in *E. coli* by J. Wang in 1971. Since then, many different but highly conserved eukaryotic and non-eukaryotic topoisomerases have been found (Gellert et al., 1976; Kato et al., 1990; Wang, 1971).

The best-known function of topoisomerases is to reduce DNA super-coiling by introducing negative coils to the double helix. They do this by creating a transient cleavage of the phosphodiester backbone of DNA followed by an under-coiling event,

and finally a trans-esterification reaction recreates the phosphodiester backbone. With a few specific variations, all topoisomerases follow the same mechanism of action. Nonetheless, these enzymes perform non-overlapping functions in living organisms. (Roca, 1995)

Given the abundance of these enzymes in nature, they have been classified into three subgroups based on their DNA cleavage ability: Type I-5' DNA topoisomerases, Type I-3' DNA topoisomerases, and Type II DNA topoisomerases (Roca, 1995). The first group, Type I-5' DNA topoisomerases, include *E. coli* DNA Topo I, Topoisomerase III (Topo III) and *S. cerevisiae* DNA Topo III, among others. These bind and cleave single-stranded DNA, work as monomers and are ATP-independent. They are involved in relaxation of negatively supercoiled DNA, knotting of single-stranded DNA rings, and joining of single-stranded rings into double-stranded rings. The second group, type I-3' DNA topoisomerases, include all eukaryotic and some viral Topo I enzymes. In this group the enzymes are monomeric ATP-independent molecules that, unlike Type I-3' DNA topoisomerases, bind double-stranded DNA and, like them, cleave only one DNA strand. They can relax both over-wound and under-wound DNA. Type II DNA topoisomerases, the last of the subgroups and the one this work focuses on, includes all Topo II in all living organisms. These are highly conserved dimeric enzymes that require ATP. They eliminate DNA super-coils or interconvert DNA knots with unknotted forms if dealing with one DNA molecule, and catenate/decatenate DNA if dealing with two different polymers. ATP is needed for proper function of this group of enzymes but it has

been shown that, in the presence of non-hydrolysable ATP, Topo II conserves its ability to bind DNA and to stabilize DNA crossovers or knotted DNA segments (Roca, 1993).

Topo II is associated with several cellular processes such as chromosome segregation, condensation, homologous recombination, DNA replication and transcription, among others (Bakshi et al., 2001). The enzymatic activity of Topo II may be involved in mitotic and meiotic chromosome segregation as an aid in the resolution of inter- and intra- chromosomal tangles that are produced when condensed chromosomes come together and/or separate (Bermejo et al., 2007; Rose et al., 1990). Specifically, Topo II functions to relieve the tangling experienced by sister chromatids at precatenate nodes. These nodes are chromosomal intertwines formed by a rotational event at the replication fork branching point and are the product of an attempt to relieve negative super-coiling caused by chromosomal handling during biological processes (Postow et al., 2001; Wang, 2002).

During transcription, RNA Polymerase II (RNA Pol II) causes dramatic changes to DNA topology, requiring Topo II to relieve the strain caused by this process (Koster et al.). Originally, the role of Topo II in transcriptional regulation was thought to be limited to the resolution of problems created by the movement of RNA Pol II as it transcribes DNA into RNA (Haince, 2006). However, it is now known that Topo II is also recruited to regions of regulated gene transcription to produce a nucleosome-specific DNA double-strand break (Bong-Gun Ju, 2006). Presumably this break is needed for transcriptional activation.

Throughout replication, the movement of a replication fork creates super-coiling. Not relieving this strain could eventually lead to unbearable topological stress that would result either in fork collapse or in breakage of the double-stranded helix (Bermejo et al., 2007). Thus, Topo II plays an essential role in replication by producing the phosphodiester breakage and rejoining necessary to relieve the strain. Topo II has been shown to be in proximity to replication forks. In mutants of Topo II the DNA damage checkpoint is activated, suggesting that the presence of Topo II during replication, and presumably its catalytic actions, are needed for proper progress through the cell cycle (Bermejo et al., 2007). However, the involvement of Topo II in replication is not limited to the elongation step (fork progression). Topo II has also been associated with initiation and termination (Baxter and Diffley, 2008; Cuvier et al., 2008; Fachinetti et al., 2010; Halmer et al., 1998; Ishimi et al., 1992).

The biological functions of Topo II extend beyond those described into overall maintenance of chromosome structure. Substantial data suggest that Topo II plays a key role in chromosomal scaffolding, the attachment of DNA to an insoluble matrix in the nucleus through which higher-order chromatin structures are achieved (Bakshi et al., 2001). Topo II is one of the main components of the nuclear matrix scaffold and is thought to form the molecular backbone of chromosomes (Earnshaw et al., 1985). Further, Topo II localizes to matrix-associated regions of DNA. These are the regions through which DNA loops attach to the chromosome scaffold (Adachi et al., 1989). Interestingly, the role it might play in the formation of higher order chromatin structure

could be uncoupled from its enzymatic action, uncovering a whole new mechanism of action for Topo II (Bojanowski et al., 1998). Although in the absence of ATP Topo II is not able to mediate relaxation of super-coiled DNA, it is able to clamp and hold two DNA strands, suggesting that its DNA-binding ability is independent from the ATP-driven conformational change.

From all the data mentioned it is evident that Topo II is an enzyme that has evolved to fulfill multiple functions in living organisms. These roles are fulfilled either through its enzymatic actions or through its DNA-binding ability. Although much has been done to understand the function of this protein most studies to date have focused on the enzymatic role of Topo II. However, in order to understand how the many functions mentioned are carried out, it is necessary to examine its global distribution and the mechanism by which its localization is controlled.

Here, I use Chromatin Immunoprecipitation (ChIP) followed by microarray analysis (ChIP-chip) to determine the global distribution of Topo II across the *Drosophila melanogaster* genome. The results suggest that, rather than associating with discrete sites in the genome, Topo II binds to large DNA regions, potentially driven in a DNA context-specific manner instead of a sequence-specific fashion. Additionally, data suggest an interplay between insulator proteins and Topo II. Chromatin insulators are classically defined as protein-DNA complexes that are thought to ensure proper attainment of a transcriptional program by (a) controlling the interactions between enhancers and

promoters and by (b) creating boundaries that prohibit inappropriate interactions between adjacent chromatin domains (Bushey et al., 2008; Ong, 2009).

In *Drosophila* there are 5 known insulators characterized by their DNA-binding proteins: Suppressor of Hairy-wing (Su(Hw)), Boundary element-associated factor (BEAF), *Drosophila* CTCF (ditch), Zeste-white 5(Zw5) and GAGA-associated factor (GAF) (Maeda and Karch, 2007). Some of these insulators have been shown to contact one another organizing the genome by mediating intra- and inter-chromosomal interactions that create topologically constrained domains (Gerasimova and Corces, 1998; Majumder and Cai, 2003; Valenzuela and Kamakaka, 2006). Specifically, two of the insulator complexes that are thought to interacts are Su(Hw) and dCTCF (Gerasimova and Corces, 1998). The interaction among these insulators is facilitated by the common use of the Centrosomal Protein 190 (CP190). Many insulator complexes are common at specific loci for different cell types throughout different stages of development. Nonetheless, this is not always the case. Different insulator subclasses may have specialized functions (Gurudatta and Corces, 2009). Particularly, some insulators show differential localization with respect to genomic landmarks, levels of gene expression, and/or association with genes involved in different cellular processes. Although the majority of the insulator proteins are found at intergenic regions a significant portion of them localize to gene-rich segments, both at introns and exons. Analysis of the different insulator subclasses at these sites reveals that Su(Hw) is mainly found in introns and is associated with genes of low expression levels. It also shows that BEAF is found mainly in exons and is skewed towards the 5' untranslated region (UTR) of highly expressed

genes associated with metabolic processes. Finally, dCTCF is found in lower but still relevant proportions than BEAF at introns also skewed towards the 5' UTR of highly expressed genes associated with developmental processes (Bushey et al., 2009). These findings have lead to the idea that different insulators have evolved divergent roles in chromatin organization and gene regulation.

Interestingly, Topo II seems to be related to insulators. Experiments by Nabirochkin et al (Nabirochkin et al., 1998) demonstrate that Topo II is localized to the *gypsy* insulator (bound by Su(Hw)) and in genetic experiments measuring insulator function, loss of Topo II results in a reversion of the insulation mutation (Ramos et al., 2011) .Together these results suggest that Topo II may be involved as a regulator of insulator function or Topo II itself is directed by insulators. Here I present data that suggest the former by indicating that insulator complexes delimit the localization of Topo II. Further, I show that Topo II behaves differently around different insulators, and that disruption of specific insulator proteins leads to a change in the distribution of Topo II around the insulator binding sites.

# Materials and Methods

## Antibodies

Rabbit α-Topo II antibody obtained from Paul Fisher at the Department of Pharmacological Sciences, State University of New York, Stony Brook was used for all Topo II immunoprecipitation (IP) experiments. Specifically, the antibody was raised in rabbits against amino acids 534-950 of Topo II. Rabbit α- Centrosomal Protein 190 (CP190) was prepared as described (Pai et al., 2004).

## ChIP-chip analysis

Chromatin immunoprecipitation (ChIP) was carried out with $3 \times 10^7$ to $5 \times 10^7$ *D. melanogaster* Kc cells at 80% confluency. Cells were cross-linked with 1% formaldehyde for 10 minutes at room temperature. Nuclear lysates were sonicated 12 times for 10-second intervals to generate 200-1000 bp DNA fragments. All samples were pre-cleared with GE Healthcare rProtein A Sepharose Fast Flow beads. Immunoprecipitation (IP) was then performed with 6 μL of rabbit α-Topo II antibody overnight at 4 C and the pull-down was done with said beads (50ul beads/sample) after a 2-hour incubation period at 4 C. Samples were washed with low-salt, high-salt, and LiCl buffers. The DNA was eluted with IP elution buffer and the DNA was obtained through Phenol:Chloroform extraction. For microarray analysis (ChIP-chip), samples were amplified two times using the GenomePlex Complete Whole Genome Amplification kit (Sigma, WGA2) to obtain a sufficient amount of DNA. Sample labeling, hybridization, and peak analysis was then performed by NimbleGen using 2.1 M whole-genome tiling arrays. Two biological

replicates were done and quantile normalization was performed between the biological replicates by averaging the height of two enriched regions. NimbleGen Signal Map was used to visualize the data. This process was repeated with two biological replicates in which Kc cells were treated with double-stranded RNA (dsRNA) corresponding to the CP190 gene in order to downregulate the expression of this gene by RNA interference (RNAi).

## Real-time PCR analysis

Real-time Polymerase Chain Reaction (PCR) analysis for random peak validation was performed with the ChIP samples that were used for microarray analysis. Fermentas Maxima SYBR Green/ROX qPCR Master Mix (#K0223) was used and ChIP enrichment was calculated by comparison of DNA concentration to a three-point standard curve from the input sample after PCR amplification.

## RNA interference

Small-interfering RNA (siRNA) -mediated knock down of CP190 was carried out using 3 different sets of primers to amplify sections from the second exon of CP190 that would then serve as the interfering double-stranded RNA (dsRNA). *LacZ* dsRNA was used as a control. The dsRNA was delivered on day 1 to Kc cells at a concentration of 4 µg/ml along with 8 µl/ml of Invitrogen's Cellfectin II reagent. This treatment was repeated for four consecutive days and the cells where then extracted on the $5^{th}$ day for

ChIP. Downregulation of CP190 protein expression was confirmed by western blot analysis using the rabbit α-CP190 antibody described above and an appropriate rabbit secondary antibody.

## Peak Analysis

Regional enrichment was determined through CMARRT (Kuan et al., 2008), using a 500 bp sliding window and a stringency of 1% false discovery rate (fdr). Anchoring of CP190 and analysis of Topo II was done using R with an algorithm developed by members of the Corces Lab. Such algorithm follows similar statistical validity as that observed in several published genomic studies where high-throughput techniques are used. CP190 subgroups were generated by using CP190, BEAF, CTCF, and Su(Hw) ChIP-chip data previously published by the Corces lab (Bushey et al., 2009) and the GAGA ChIP-chip data obtained from the modENCODE project website (http://www.modencode.org). Galaxy (http://main.g2.bx.psu.edu) was then used to subgroup these datasets.

## Western Blot Analysis and Immunoprecipitations

For western analysis, Kc cells were prepared using standard protocols and run on tris-glycine gel as describe in Gerasimova *et al.* using the SDS sample buffer (Gerasimova et al., 2007). For all the protein immunodetections the Millipore SNAP i.d. protein detection system was used following the manufacturers protocol. Blots were probed with mouse α-Lamin at 1:5000 and rat α-CP190 at 1:10000.

# Results

## Topo II binds to genomic intervals

In order to study the localization of Topo II on a genome-wide scale I performed a ChIP-chip. Figure 1 shows the raw distribution of Topo II across a 500 Kb segment of chromosome 2L. It is evident that Topo II does not bind narrow sites in the genome but rather binds to broad genomic intervals, which I define as broad segments in the linear genome with wide deviation from the center of the enriched region as opposed as narrow sites.

To determine the specificity of the microarray analysis I chose 10 random loci where Topo II was either highly enriched or not enriched and developed primers to verify these sites through real-time PCR (Table 1). With this method I verified 10 out of the 10 sites. Comparison of the sites where Topo II binds in high proportion to those where it does not shows an average six-fold difference. This result suggests the absence of large number of false positives in the data.

Next, Topo II enrichment was identified. Using traditional peak calling method widely utilized for identifying transcription factors or insulator binding sites failed when analyzing the Topo II datasets. These programs are limited to the identification of very refined and defined enrichment. Thus, in order to highlight the regions were Topo II is enriched I used CMARTT, a tool that uses an averaging-determining moving window to determine enrichment over wide genomic regions rather than over short genomic probes. In addition, CMARRT accounts for the correlation between the measurements from probes that map to consecutive genomic locations, which based on the array design, may

signal a false enrichment (Kuan et al., 2008). Figure 2 shows a 1.5 Mb region of chromosome 2L where the purple track reflects the binding of Topo II and the green track shows the enriched regions as determined by CMARRT. I used a 500-bp sliding window and a 1% fdr to determine enriched regions. Comparison of the purple and green tracks along the whole genome highlights enrichment. However, even with CMARRT, a true representation of the enrichment visualized in the raw data is one of the shortcomings for peak calling with proteins that span over large genomic intervals or over gene bodies, mainly because significant amount of data that does not make it past the stringent cut-off is lost. Thus, for visualization purposes, CMARRT was used to call peaks but the raw data files were used for all other analysis. These results make evident the fact that Topo II is found across much of the genome. Nonetheless, rather than showing uniform, unspecific binding or narrow localization regions, it associated with wide genomic areas.

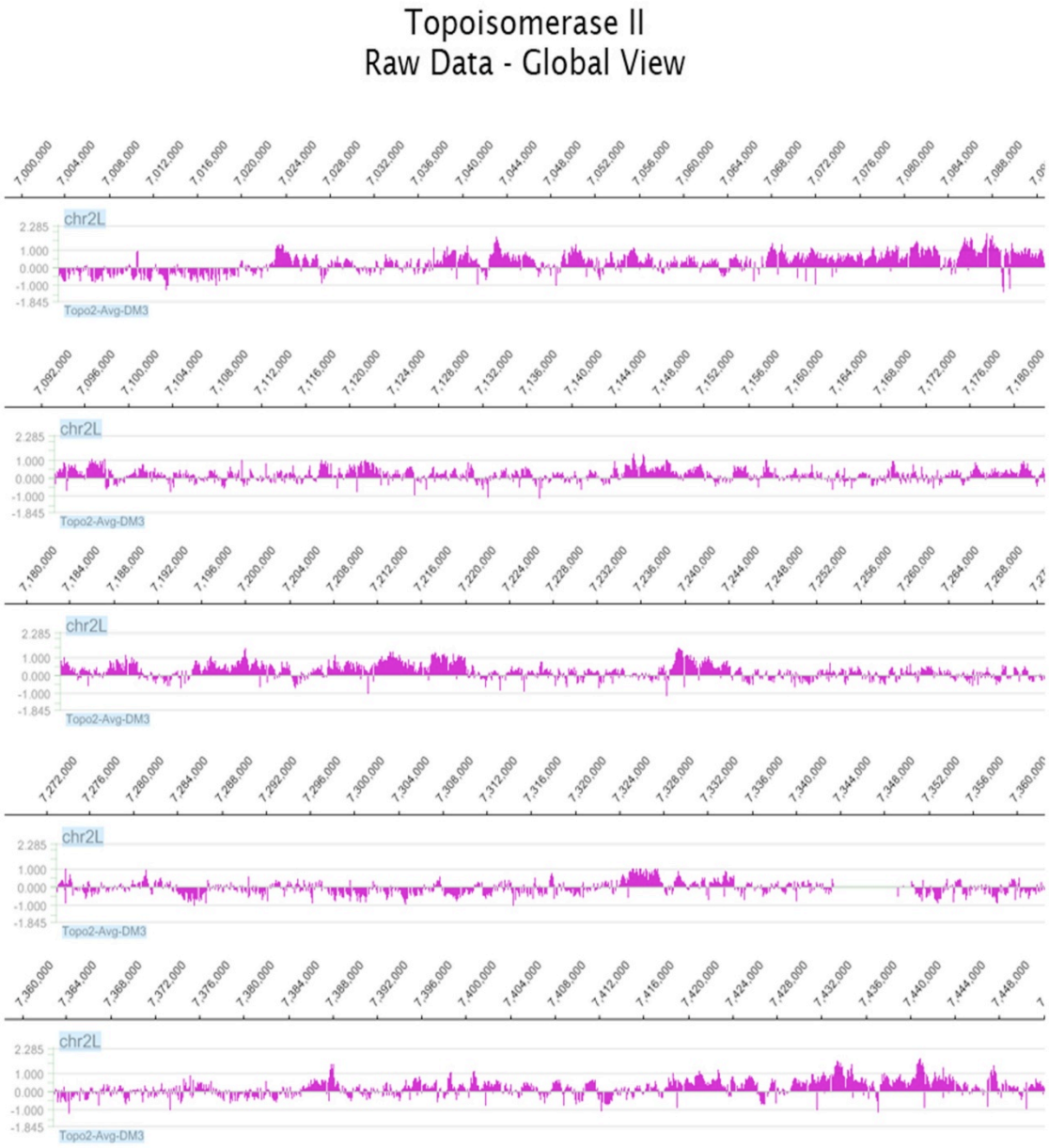



**Figure 1 – Genome-wide localization of Topo II**

Representative ChIP-chip data for Topo II in a 500 Kb region of chromosome 2L. The Y-axis expresses the log2 ratio of ChIP/Input signal. The X-axis represents the linear genome. These data are the result of two biological replicas.

| Chromosome | Start | End | Enriched Region | ChIP Enrichment | No Ab Enrichment |
|---|---|---|---|---|---|
| chr2L | 125000 | 126000 | Yes | 0.021795 | >4E-08 |
| chr2L | 7813000 | 7814000 | Yes | 0.025598 | 2.59E-04 |
| chr2L | 13576000 | 13577000 | Yes | 0.011227 | >4E-08 |
| chr3R | 1132000 | 1133000 | Yes | 1.99E-02 | 1.49E-04 |
| chr3R | 17422000 | 17423000 | Yes | 7.52E-01 | 0.00165 |
| chr2R | 19603000 | 19604000 | Yes | 0.101466 | 3.41E-04 |
| chr2R | 14321000 | 14322000 | Yes | 0.010461 | 1.34E-06 |
| chr2R | 9445000 | 9446000 | No | 5.64E-03 | 1.59E-04 |
| chr2L | 20045000 | 20046000 | No | 1.29E-03 | 3.73E-08 |
| chr2L | 5401000 | 5402000 | No | 7.32E-03 | 5.05E-04 |

**Table 1 - qPCR verification of ChIP-chip**

For each one of the ten sets of primers used for ChIP verification I show the start and end coordinates, whether this is a Topo II enriched region according to the microarray analysis, and the enrichment levels relative to a 3-point standard curve for both ChIP sample and no-antibody sample (No Ab). This table shows that for all cases there is an enrichment of the ChIP over the No Ab control, and that regions enriched according to the microarray analysis show significant enrichment over those regions where TopoII is predicted to be at low levels according to the ChIP-chip.

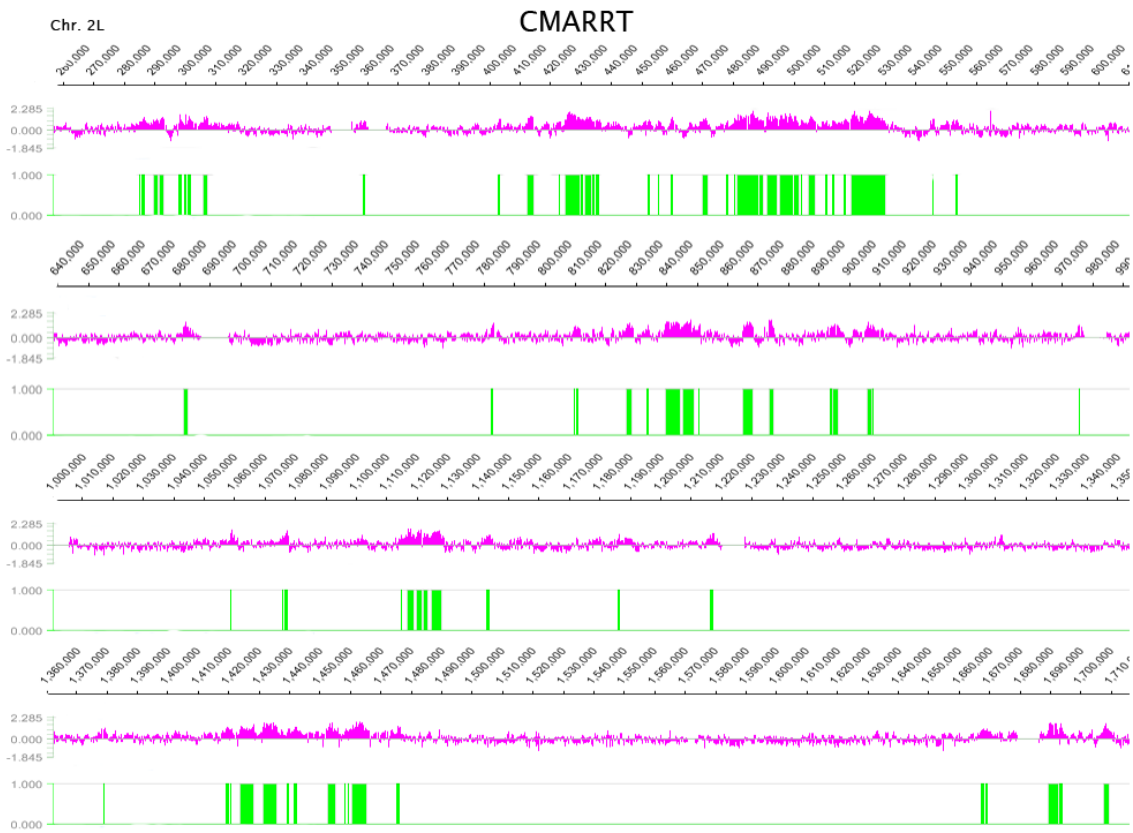



**Figure 2 - Statistical analysis of Topo II raw data using CMARRT**

Shown is a 1.5 Mb region of chromosome 2L representative of the analysis done using CMARRT. The purple tracks represent the raw data for Topo II ChIP-chip as in figure 1. The green tracks are enriched regions as determined by an average-determining window with a 500 bp span and a 1% fdr. In this analysis a binomial system is used where only values of 0 or 1 are allowed as represented by the Y-axis of the green tracks.

## Topo II distribution is flanked by CP190

Analysis of the global distribution of Topo II across the genome revealed that although this enzyme is present at low levels across much of the genome and at high levels across several genomic intervals, its distribution is not random and is flanked by regulatory elements known as insulator proteins.

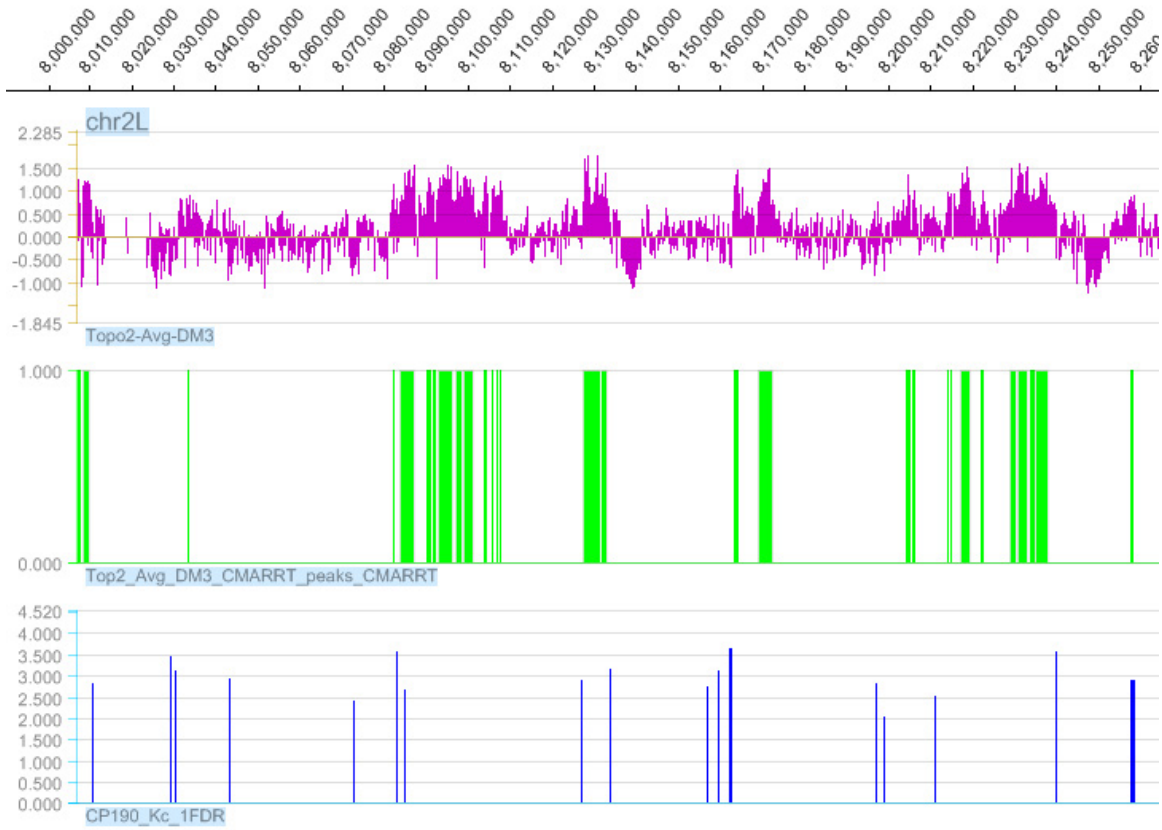
I overlapped the ChIP-chip data obtained experimentally for Topo II with that of CP190. CP190 has been shown to interact with other insulator proteins and is essential for insulator function in *D. melanogaster* (Gurudatta and Corces, 2009). Figure 3A represents a 200 Kb region of chromosome 2L. As in previous images, in purple and green are the raw data for Topo II ChIP-chip and CMARRT analysis respectively. In blue is a ChIP-chip data set for CP190 (Bushey et al., 2009).

After careful analysis I noticed that CP190 seems to flank many of the segments where Topo II is notably enriched. Further, it seems to interrupt the binding of TopoII where this last one is excluded from the CP190 insulator sites. Vertical tracking of CP190 peaks over either CMARRT data or raw data of Topo II makes this fact evident.

To verify this possibility computationally, I anchored all the CP190 sites and extracted the enrichment values for Topo II over a 2 Kb window using the raw data file (Figure 3B). Anchoring all known CP190 peaks shows a decreased enrichment of Topo II as it approaches the center point for CP190, followed by a return to enrichment levels

similar to those observed before the CP190 anchor. In other words, CP190 seems to interrupt the enrichment of Topo II such that Topo II makes a "dip" in its distribution at sites where CP190 is present. The Cohesin complex, which is involved in sister chromatid cohesion during replication, shows an overall distribution surprisingly similar to that of Topo II as evidenced by ChIP-chip studies and polytene chromosomes (Figure 4). However, as shown in Figure 4D, Cohesin seems to be unaffected by the presence of CP190 and may even have a preference for binding with CP190 at these sites. Together these observations help support the idea that the distribution of Topo II may be delimited by the insulator protein CP190, or at least by a complex to which CP190 associates.

A




B

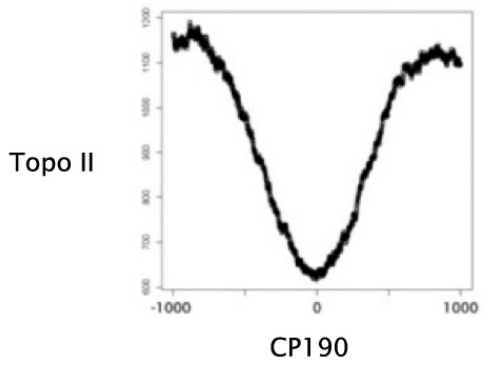



**Figure 3 - TopoII is flanked and interrupted by CP190**

**(A)** Representative 250 Kb segment of chromosome 2L showing ChIP-chip data for both Topo II and CP190. In purple and green are raw data and CMARRT analysis for Topo II respectively. In blue is a 1% fdr ChIP-chip dataset for CP190. The Y-axis of Topo II raw data and CP190 express the log2 ratio of ChIP/input signal. The Y-axis for CMARRT

Topo II analysis expresses a regional enrichment of Topo II binding where only levels of 0 - no enrichment - and 1 - enrichment - are allowed. Tracking of a high portion of CP190 peaks in a vertical fashion over Topo II enriched regions shows that CP190 colocalizes to the genomic regions to which Topo II binds but rather than overlapping with it, it seems to flank it. **(B)** Global image of the distribution of Topo II around CP190 peaks. The Y-axis expresses the sum of the average levels of enrichment of Topo II based on a log2-scale ratio. The X-axis is a 2 Kb window of the linear distance around all known CP190 peaks. The 0-mark represents the anchoring point where all CP190 peaks are taken to be.

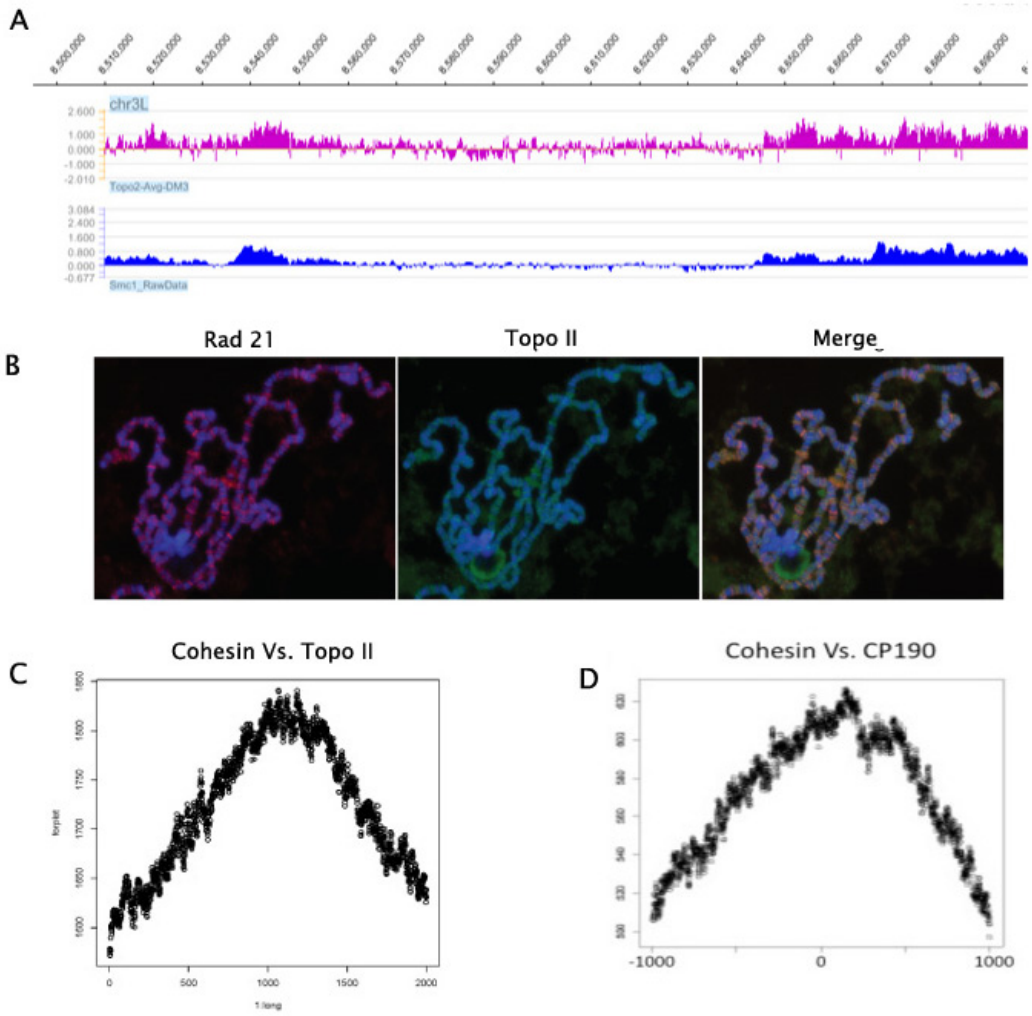



**Figure 4 - Cohesin localization around Topo II and CP190**

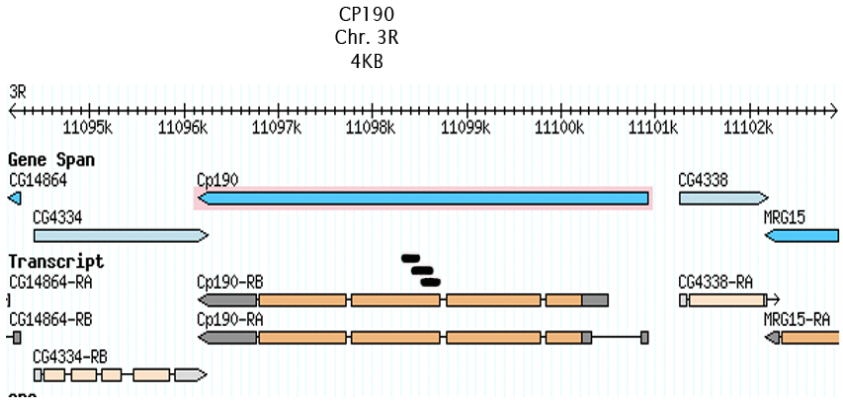
The Cohesin complex, which has a distribution similar to that of TopoII is unaffected by CP190. **(A)** Representative 200 Kb segment of chromosome 2L showing ChIP-chip data for both Topo II (purple track) and Smc1 (a member of the Cohesin complex) (blue track). This image shows a surprising similarity in the localization pattern between TopoII and Smc1. **(B)** *Drosophila* polytene chromosome immunostaining for Rad 21 (a member of the Cohesin complex) and Topo II supports the idea that these two proteins colocalize across the genome. **(C and D)** Global image of the distribution of Cohesin around Topo II and Cohesin Around CP190, respectively. The Y-axis expresses the sum of the average levels of enrichment of Topo II based on a log2-scale ratio. The X-axis is a 2 Kb window of the linear distance around all known Topo II (C) and CP190 (D) peaks where the 0-mark represents the anchoring point where all peaks are taken to be.

## Loss of CP190 leads to a change in Topo II distribution

To study whether there is functional relationship between CP190 and Topo II, I knocked down CP190 in Kc cells using RNAi and then used these cells to perform a ChIP-chip experiment using antibodies against Topo II (Figures 5-7).

To further study the relationship between CP190 and Topo II, I graphed the overall distribution of Topo II after CP190 knock down around the known CP190 binding sites (Figure 7). Comparison of this analysis with that on Figure 3B shows that where formerly Topo II was found at low levels of binding over CP190 genomic sites in wild type cells, it is enriched after CP190 knock down. These results support the idea that CP190 and Topo II are related in such a way that CP190, or an element associated with it, delimits the binding of TopoII and without CP190 to serve as a barrier Topo II is free to associate with those genomic regions.

A




B

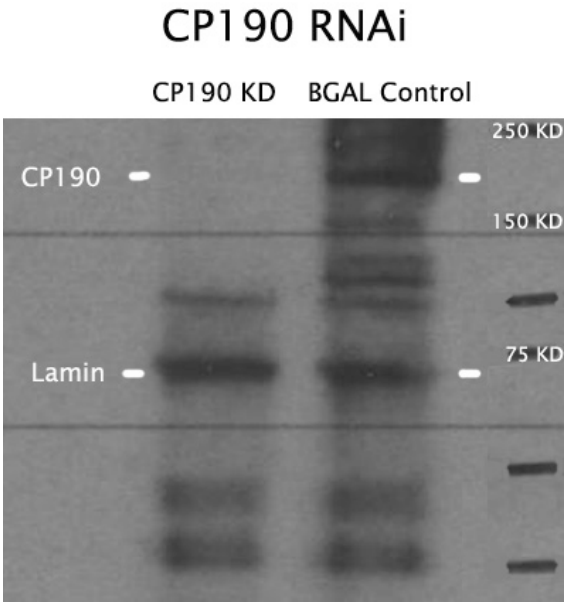



**Figure 5 - CP190 knock down**

**(A)** Gene and transcript map of CP190 obtained and modified from flybase. In dark blue is the gene segment in chromosome 3R. In yellow are the CP190 transcripts represented in exon regions. All 3 sets of primers I used map to the second exon as shown by the thick, black horizontal lines above the exon. **(B)** Western confirming successful CP190 knock down. In the left is the CP190 knock down sample and in the right a β-Gal control. Lamin is used as loading control. Molecular weights in kDa: CP190 – 190; Lamin – 70.

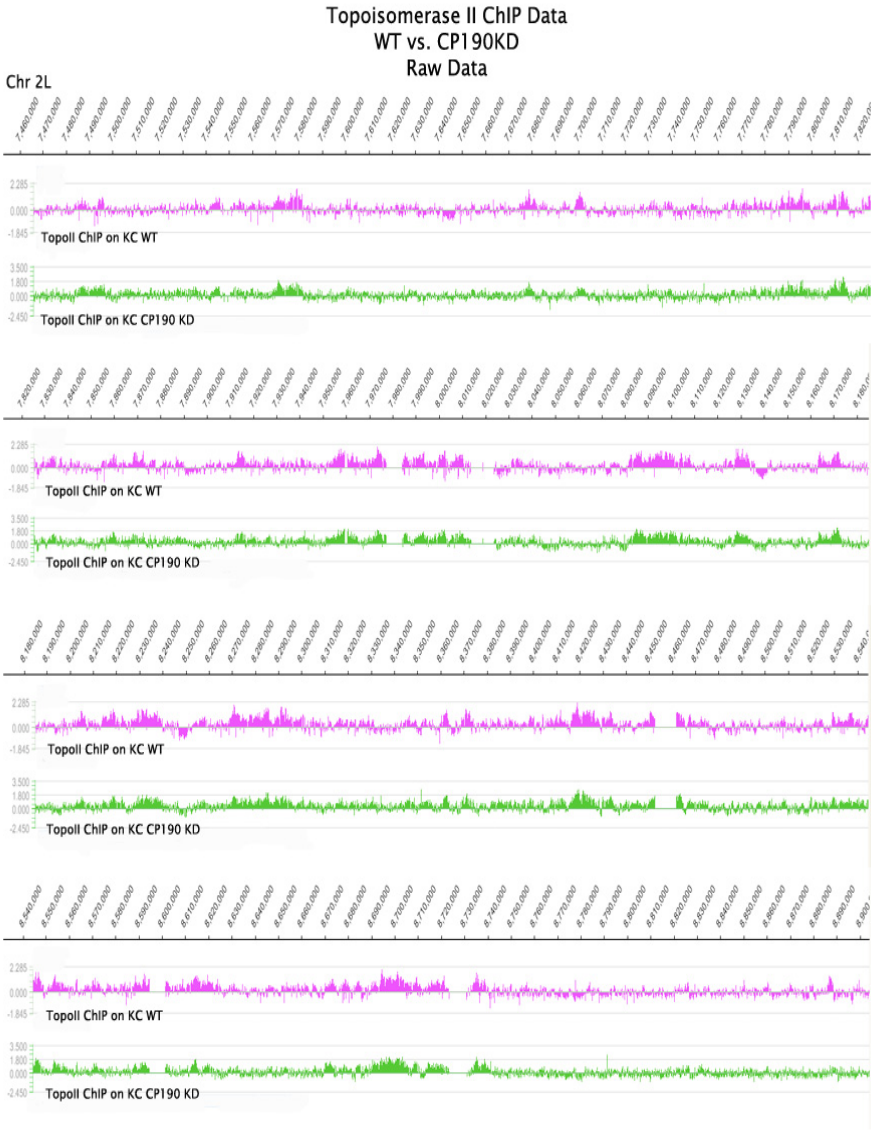



**Figure 6 - Topo II ChIP-chip on CP190 knock down**

A 1.5 Mb region of chromosome 2L representative of the pattern observed throughout the whole genome. In purple we find Topo II raw ChIP- chip data in wild type cells. In green is the data for Topo II ChIP-chip after CP190 RNAi. Both ChIP's where done with Kc cells. I used the same protocol for both datasets and other than the RNAi treatment all conditions were identical. The distribution of Topo II across genomic segments is roughly identical in both wild type and CP190 knock down when observed at low resolution.

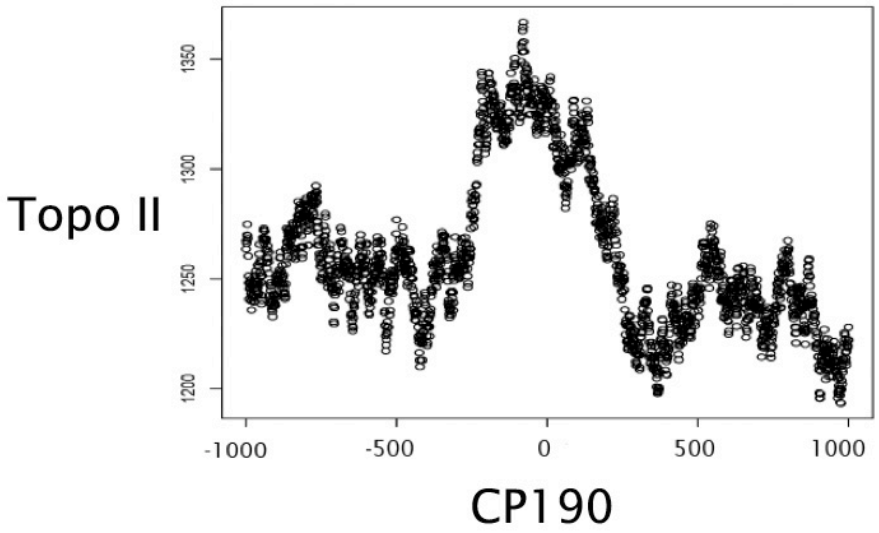



**Figure 7 - Distribution of Topo II around CP190 after CP190 RNAi**

This represents the global distribution of Topo II around CP190 binding sites where the Y-axis expresses the sum of the average levels of enrichment of Topo II based on a log2-scale ratio. The X-axis is a 2 Kb window of the linear distance around all known CP190 binding sites below a 1% fdr threshold. The 0-mark represents the anchoring point where all CP190 peaks are taken to be.

## The behavior of Topo II at CP190 sites is not the same in all insulator subgroups

CP190 is an insulator protein present in four of the five known *Drosophila* insulator complexes: Su(Hw)-, dCTCF-, BEAF-, and GAF- (Gurudatta and Corces, 2009). However, each of these DNA-binding proteins has a distinct distribution in the genome. Thus, given the possibility that CP190 may delimit Topo II localization I analyzed the behavior of Topo II in the different CP190 insulators. I generated 5 different datasets (Figure 8) representative of those CP190 regions that associate only with Su(Hw) (1365 sites), dCTCF (604 sites), BEAF (1556 sites), GAF (126 sites), and those CP190 regions that do not associate with any of the known insulator proteins (771 sites).

To study the behavior of Topo II around these CP190 subgroups I repeated the analysis done previously where CP190 is anchored and Topo II enrichment is observed in a 2 Kb surrounding window (Figure 9). When looking at those sites where CP190 is either by itself or with an unidentified protein it appears that this insulator protein provides a barrier for Topo II in wild type cells. This barrier is lost when CP190 is knocked down, which presumably allows for an enrichment of Topo II. This same pattern is observed for those CP190 sites that contain CTCF, BEAF or GAF. Notably, in the subgroup where CP190 interacts with GAF, loss of CP190 has a lesser effect on the distribution of Topo II at the anchoring point as compared to CTCF- and BEAF-containing subgroups. The remainder of a "dip" at the anchoring point shows this. Interestingly, the arrangement of Topo II around those CP190 sites shared with Su(Hw) differs substantially from that observed in other subgroups. In both wild type and CP190

knock down Kc cells Topo II appears to be enriched at the anchoring point. In these cases the presence of CP190 seems to have little relevance, if any, to the enrichment and distribution of Topo II.

Topo II appears to associate to wide genomic intervals evidenced by its binding to broad loci. The multiple functionality attributed to Topo II suggest that there must exist a mechanism that regulates its localization. In order to function properly it has to be spatial-temporally regulated. My data suggests that, at least in part, this may be accomplished by a CP190 barrier as observed by the reduced levels of Topo II around CP190 and later increase over those regions in CP190 knocked down cells. This distribution is conserved when looking at the different insulator subgroups with the exception of the Su(Hw) insulator.

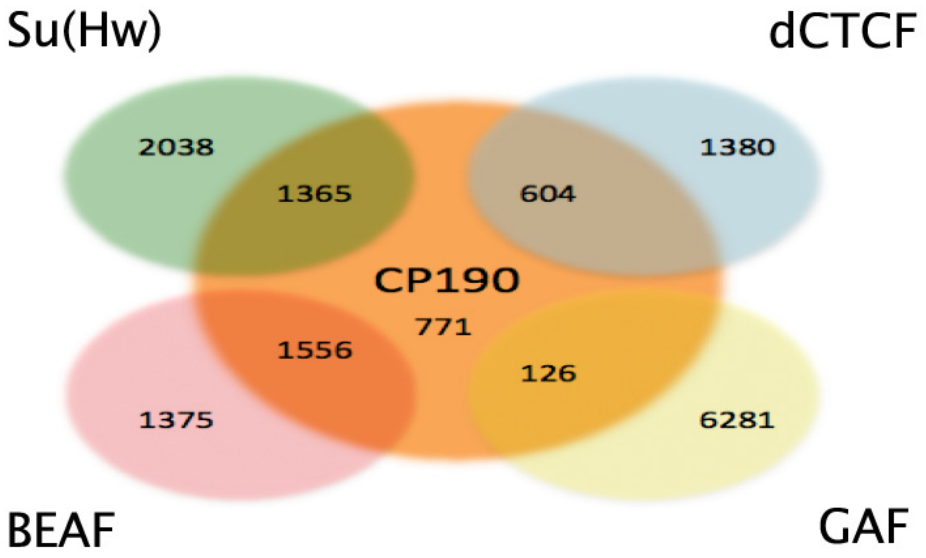



**Figure 8 - Venn diagram of the different CP190 subgroups**

These subgroups were generated to study the behavior of Topo II around different insulators. Five different CP190 subgroups were generated: CP190 alone (771 sites), CP190 and Su(Hw) (1365 sites), CP190 and dCTCF (604 sites), CP190 and GAF (126 sites), and CP190 and BEAF (1556 sites). To generate these groups I used experimentally determined binding sites for these proteins by the Corces lab. Additionally, I subtracted those CP190 sites that localize to more than one insulator protein. As a result about 10% of the CP190 binding sites are left out of this study.

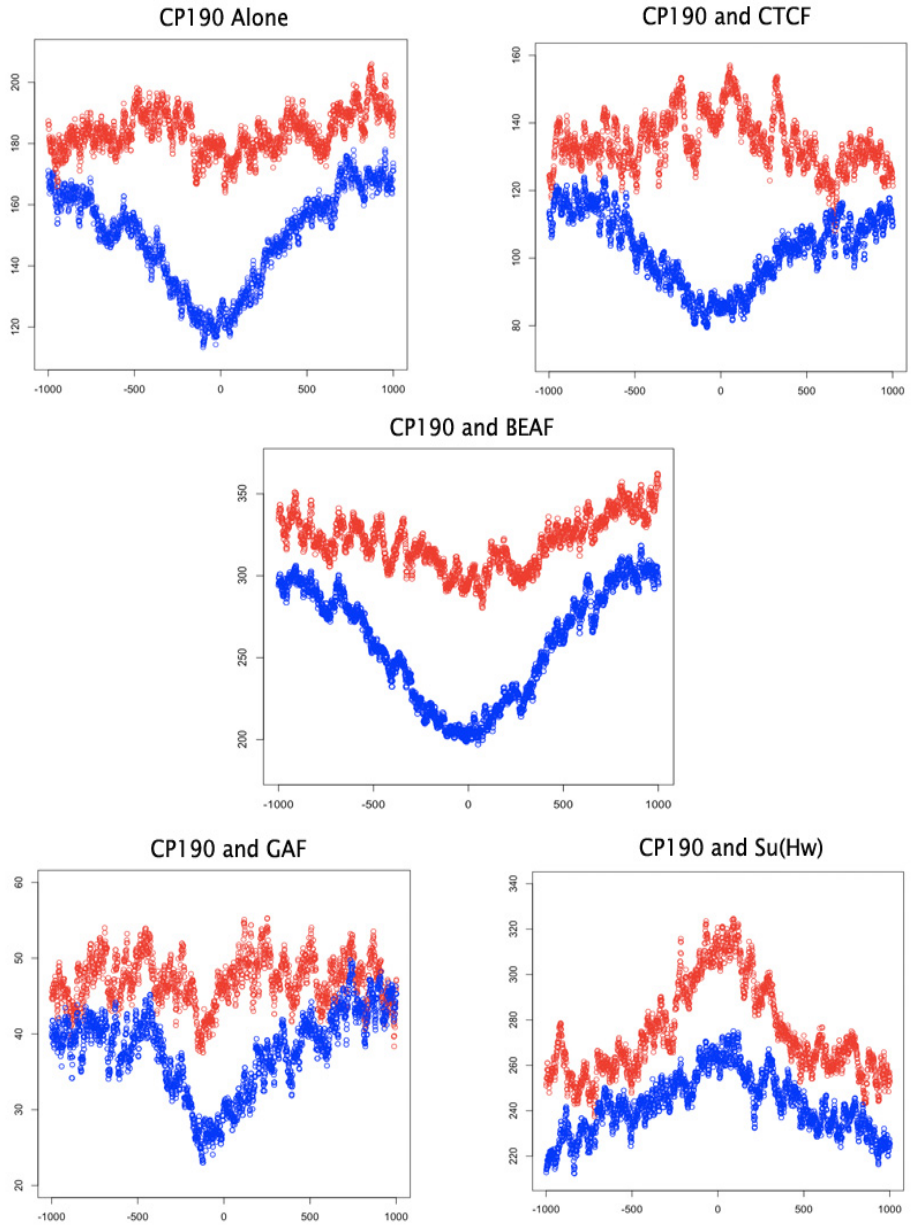



**Figure 9 - Topo II distribution around different CP190 subgroups**

Topo II distribution around different CP190 subgroups in wild type (blue) and CP190 knocked down cells (red). For all plots, the Y-axis expresses the sum of the average levels of enrichment of Topo II based on a log2-scale ratio and the X-axis represents a 2 Kb window of the linear distance around the CP190 binding sites below a 1% fdr threshold that correspond to a specific subgroup of insulators. CP190 seems to act as a barrier for Topo II enrichment at those places where it interacts with CTCF, BEAF, GAF, or is present by itself (or interacts with an unidentified complex). In these same subgroups, loss of CP190 affects the arrangement of Topo II across the anchoring point. The Topo II binding pattern across the CP190 binding segments that interact with

Su(Hw) is different to that observed previously. In these, rather than “dipping” at the anchoring point, Topo II appears to be enriched and loss of CP190 seems to have no significant effect on Topo II. It is worth mentioning that the difference in the Y-axis label between graphs is a product of the number of binding sites present in each subgroup.

# Discussion

Topo II is an essential component in a multitude of cellular processes. However, its global binding pattern throughout the genome has not been extensively studied. Using *Drosophila* polytene chromosomes, I have shown through immunofluorescence that Topo II's distribution is not uniform as has been proposed in the past (Bakshi et al., 2001) but is rather enriched at multiple, specific loci (Ramos et al., 2011). Topo II is localized to euchromatic regions, possibly working in areas were transcription is occurring. In addition, ChIP-chip experiments support these results by showing that rather than equally binding throughout DNA Topo II is enriched at wide genomic areas, many of which localize with active chromatin marks (data not shown).

Topo II has been shown to weakly bind DNA even in the absence of ATP (Hizume et al., 2007). In a titration experiment the presence of this enzyme on DNA increased as its concentration increased. This suggests that Topo II might not have high binding specificity. Here I report two types of binding patterns for Topo II. On the first pattern Topo II is enriched at large but particular loci. In the second, this enzyme shows low-specificity binding to large portions of the genome. This last type of binding pattern could be explained by the presence of biologically irrelevant background isolated throughout the procedure. Nonetheless, this is an unlikely possibility due to careful analysis of the data that shows that much of the low-level binding is consistent in biological replicas. The other possibility, and a more likely one, is that Topo II has little, if any, sequence specificity. If this is the case though, what can localize Topo II to the multitude of loci where it is highly enriched? Comparison of the peak analysis done in this work with the expression profile of Kc cells shows that Topo II is mainly found in

regions that contain active genes (data not shown). Throughout transcription there is a change in chromatin structure that allows for proteins involved in this process to access DNA (Hizume et al., 2007). Thus, Topo II could localize to these regions in a structure- or context-specific manner. Recently this has been shown to be true as Topo II requires a topological change on the DNA to drive its localization and decatenation function (Baxter et al., 2011; Kegel et al., 2011).

Along with multitasking capability comes the necessity of proper regulation. As it has been alluded to, Topo II is a multifunctional enzyme. Replication and transcription are not the only processes in which it is involved. Different studies have related it to DNA condensation (Rose et al., 1990), chromosome segregation (Lupo et al., 2001), recombination (Christman et al., 1988), formation of higher-order chromatin structure (Hizume et al., 2007), and even chromatin looping through chromosome scaffolding (Adachi et al., 1989). Involvement in such a variety of processes requires a mechanism that provides spatial-temporal direction for Topo II. So far no such process has been proposed. Here I make a first and crucial step towards the discovery of such a process by showing that CP190, an insulator protein, provides a boundary for the localization of Topo II.

With the exception of Zw5, all known *Drosophila* insulators interact with CP190; this interaction is required for their proper functioning (Gurudatta and Corces, 2009). This fact and the pattern observed between CP190 and Topo II suggest that a further way in which insulator complexes may affect genome organization is by regulating the

localization of Topo II. Presumably, these complexes maintain Topo II within particular chromatin domains or may provide anchoring points in the genome that then allow for the twisting and unwinding of active DNA.

An argument against the functional relationship between the localization pattern of Topo II with CP190 could be that these two proteins have mutually exclusive binding motifs where the localization of CP190 flanks that of Topo II at multiple regions in the genome. This explanation introduces the possibility that Topo II and CP190 have no functional relationship and that the data present here is just an artifact. Such an explanation would predict that in the absence of CP190 there would be no change in the localization of Topo II. However, using RNAi to knock down CP190 demonstrates that Topo II's binding pattern is indeed affected. Furthermore, since CP190 is required for insulator function, this observation suggests cells may control the activity of Topo II by regulating the recruitment of CP190.

Insulator elements containing Su(Hw), dCTCF, and BEAF have been shown to have very different distribution patterns with respect to gene locations and this data also reflects differences in possible insulator function (Bushey et al., 2009). Comparing the localization of Topo II in WT and CP190 Knock down cells between each of the CP190 subgroups shows that the binding of Topo II around CP190 differs between the DNA-binding protein being utilized by CP190. For those subgroups of CP190 sites not associated with any known insulator protein or CP190 sites associated with dCTCF or BEAF dramatic changes are seen for Topo II between the WT and CP190 knock downs.

To a lesser extent Topo II localization also changes at CP190 sites associated with the GAF insulator subclasses. However, this distribution of Topo II is not significantly changed around Su(Hw) insulator binding sites when comparing WT and CP190 knock down distributions.

These differences could have functional implications for the arrangement of the chromatin fiber within the genome and the elements associated with this fiber. The 3 subclasses that appear to be most affected by CP190 are those associated with dCTCF, with BEAF, and those that are not associated with any known insulator protein. The dCTCF and BEAF insulators have been shown to interact with highly transcribed genes where they either negatively or positively affect expression (Bushey et al., 2009). Topo II's ability to relief the strain caused during transcription make it an important player at places of high transcription. Thus, cells might be using insulators as a mechanism that regulates the localization of TopoII at regions of high transcription. Additionally, since CP190 itself cannot bind DNA (Pai et al., 2004) those sites where CP190 is not associated with any of the known insulators may be sites where another yet-to-be-discovered DNA-binding protein binds and recruits CP190. Recent data suggest that those sites of CP190 alone may be sites associated with Cohesin (personal communication E. Ramos). Given the distribution similarity of TopoII around these sites with that observed at dCTCF- and BEAF- associated CP190 sites, presumably this is also an insulator associated with high level of gene expression.

While the dCTCF and the BEAF insulators are thought to regulate the expression of single specific genes, Su(Hw) insulators are believed to have a more general role in chromatin organization (Bushey et al., 2009). Around the Su(Hw)-associated CP190 sites Topo II is enriched rather than reduced. This result could be due to the possible general chromatin organization function for the Su(Hw) subclass of insulators rather than to transcriptional association associated with other insulators.

The Su(Hw) insulator requires the interaction of at least 3 proteins: Su(Hw), CP190, and Modifier of mdg4 (Mod(mdg4)2.2). If any of these element are missing the complex is not functional (Gurudatta and Corces, 2009). In previous work I showed that in Topo II mutants the binding of Topo II affects the binding of Mod(mdg4)2.2 to polytene chromosomes where the latter seems to be absent. In this same study, co-immunoprecipitation (Co-IP) of TopoII and Mod(mdg4)2.2 shows that Topo II interacts with this protein even though such an interaction is not detected on polytene stains (Ramos et al., 2011). Despite the fact that in Topo II mutants there is a slight decrease in the levels of Mod(mdg4)2.2 synthesis, most of the effects observed are due to protein degradation. Without Topo II the proteasome pathway targets Mod(mdg4)2.2 (Ramos et al., 2011). Thus, it is possible that the enrichment of Topo II at the Su(Hw) insulator complexes is explained by its stabilizing interaction with Mod(mdg4)2.2.

An argument against this explanation could be that since Mod(mdg4) is also present in the dCTCF- and GAF-insulators complexes, the same pattern should be observed in those insulators as well. However, here I show that this is the case neither for

dCTCF- nor for GAF-insulators. The existence of over 29 different isoforms of Mod(mdg4) may explain this due to the fact that different insulators may have different isoforms of Mod(mdg4) and Topo II only stabilizes the Mod(mdg4)2.2 isoform which to date is only known to bind with Su(Hw).

Creating a transitory breakage in both strands of a double-helix structure followed by an under-winding event and re-ligation of this double-helix proofs to be useful in processes like transcription, translation, condensation, segregation, and, in general, tasks in which wide chromosome movement is involved. However, given the presence of Topo II across such large genomic regions and at such high frequencies, as those observed in this study it is hard to believe that the only mechanism of action Topo II has is its well-studied ATP-dependent enzymatic reaction. The fact that Topo II can bind and clamp DNA in the absence of hydrolysable ATP suggests that Topo II also has roles independent of its catalytic activity. To understand the specific nature of such roles much more work needs to be done. Also, to fulfill the many roles suggested Topo II needs a fine regulatory system that would coordinate its functional localization. It is improbable that an enzyme involved in multiple cellular processes potentially performing different functions can fulfill its tasks without proper distribution and regulation. To elucidate a regulatory mechanism more work is needed. However, this study opens a window for the understanding of a regulatory mechanism that could account for at least part of the binding specificity that Topo II requires to do the many jobs it accomplishes. In order to further understand Topo II as a multifunctional enzyme additional studies need to look at these results in relationship to transcriptional activity, replication, and other processes in

which Topo II functions. The idea would be to study the binding pattern of Topo II around insulator proteins after partitioning its DNA association with different cellular processes and cell cycle stages. In other words, association of Topo II binding to its role in transcription, replication and other known processes could reveal novel distribution patterns for Topo II that may allow for the study of Topo II at a structural level, thus leading to a more refined understanding of chromatin dynamics.

# References

Adachi, Y., Kas, E., and Laemmli, U.K. (1989). Preferential, cooperative binding of DNA topoisomerase II to scaffold-associated regions. EMBO J *8*, 3997-4006.

Bakshi, R.P., Galande, S., and Muniyappa, K. (2001). Functional and regulatory characteristics of eukaryotic type II DNA topoisomerase. Critical Reviews in Biochemistry and Molecular Biology *36*, 1-37.

Baxter, J., and Diffley, J.F. (2008). Topoisomerase II inactivation prevents the completion of DNA replication in budding yeast. Mol Cell *30*, 790-802.

Baxter, J., Sen, N., Martinez, V.L., De Carandini, M.E.M., Schvartzman, J.B., Diffley, J.F.X., and Aragon, L. (2011). Positive Supercoiling of Mitotic DNA Drives Decatenation by Topoisomerase II in Eukaryotes. Science *331*, 1328-1332.

Bermejo, R., Doksani, Y., Capra, T., Katou, Y.M., Tanaka, H., Shirahige, K., and Foiani, M. (2007). Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation. Genes Dev *21*, 1921-1936.

Bojanowski, K., Maniotis, A.J., Plisov, S., Larsen, A.K., and Ingber, D.E. (1998). DNA topoisomerase II can drive changes in higher order chromosome architecture without enzymatically modifying DNA. J Cell Biochem *69*, 127-142.

Bong-Gun Ju, V.V.L., Valentina Perissi, Ivan Garcia-Bassets, David W. Rose, Christopher K. Glass, and Michael G. Rosenfeld (2006). A Topoisomerase IIß-Mediated dsDNA Break Required for Regulated Transcription. Science *312*, 1798.

Bushey, A.M., Dorman, E.R., and Corces, V.G. (2008). Chromatin insulators: regulatory mechanisms and epigenetic inheritance. Mol Cell *32*, 1-9.

Bushey, A.M., Ramos, E., and Corces, V.G. (2009). Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. Genes & Development *23*, 1338-1350.

Christman, M.F., Dietrich, F.S., and Fink, G.R. (1988). Mitotic recombination in the rDNA of S. cerevisiae is suppressed by the combined action of DNA topoisomerases I and II. Cell *55*, 413-425.

Cuvier, O., Stanojcic, S., Lemaitre, J.M., and Mechali, M. (2008). A topoisomerase II-dependent mechanism for resetting replicons at the S-M-phase transition. Genes Dev *22*, 860-865.

Earnshaw, W.C., Halligan, B., Cooke, C.A., Heck, M.M., and Liu, L.F. (1985). Topoisomerase II is a structural component of mitotic chromosome scaffolds. J Cell Biol *100*, 1706-1715.

Fachinetti, D., Bermejo, R., Cocito, A., Minardi, S., Katou, Y., Kanoh, Y., Shirahige, K., Azvolinsky, A., Zakian, V.A., and Foiani, M. (2010). Replication Termination at Eukaryotic Chromosomes Is Mediated by Top2 and Occurs at Genomic Loci Containing Pausing Elements. Molecular Cell *39*, 595-605.

Gellert, M., Mizuuchi, K., Odea, M.H., and Nash, H.A. (1976). DNA Gyrase - Enzyme That Introduces Superhelical Turns into DNA. Proceedings of the National Academy of Sciences of the United States of America *73*, 3872-3876.

Gerasimova, T.I., and Corces, V.G. (1998). Polycomb and trithorax group proteins mediate the function of a chromatin insulator. Cell *92*, 511-521.

Gerasimova, T.I., Lei, E.P., Bushey, A.M., and Corces, V.G. (2007). Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. Mol Cell *28*, 761-772.

Gurudatta, B.V., and Corces, V.G. (2009). Chromatin insulators: lessons from the fly. Brief Funct Genomic Proteomic *8*, 276-282.

Haince, J., Rouleau, M., Poirier, G.G. (2006). Gene Expression Needs a Break to Unwind Before Carrying On. Science *312*, 1752-1753.

Halmer, L., Vestner, B., and Gruss, C. (1998). Involvement of topoisomerases in the initiation of simian virus 40 minichromosome replication. J Biol Chem *273*, 34792-34798.

Hizume, K., Araki, S., Yoshikawa, K., and Takeyasu, K. (2007). Topoisomerase II, scaffold component, promotes chromatin compaction in vitro in a linker-histone H1-dependent manner. Nucleic Acids Res *35*, 2787-2799.

Ishimi, Y., Sugasawa, K., Hanaoka, F., Eki, T., and Hurwitz, J. (1992). Topoisomerase II plays an essential role as a swivelase in the late stage of SV40 chromosome replication in vitro. J Biol Chem *267*, 462-466.

Kato, J., Nishimura, Y., Imamura, R., Niki, H., Hiraga, S., and Suzuki, H. (1990). New Topoisomerase Essential for Chromosome Segregation in Escherichia-Coli. Cell *63*, 393-404.

Kegel, A., Betts-Lindroos, H., Kanno, T., Jeppsson, K., Strom, L., Katou, Y., Itoh, T., Shirahige, K., and Sjogren, C. (2011). Chromosome length influences replication-induced topological stress. Nature *471*, 392-+.

Koster, D.A., Crut, A., Shuman, S., Bjornsti, M.A., and Dekker, N.H. Cellular strategies for regulating DNA supercoiling: a single-molecule perspective. Cell *142*, 519-530.

Kuan, P.F., Chun, H., and Keles, S. (2008). CMARRT: a tool for the analysis of ChIP-chip data from tiling arrays by incorporating the correlation structure. Pac Symp Biocomput, 515-526.

Lupo, R., Breiling, A., Bianchi, M.E., and Orlando, V. (2001). Drosophila chromosome condensation proteins Topoisomerase II and Barren colocalize with Polycomb and maintain Fab-7 PRE silencing. Mol Cell *7*, 127-136.

Maeda, R.K., and Karch, F. (2007). Making connections: boundaries and insulators in Drosophila. Curr Opin Genet Dev *17*, 394-399.

Majumder, P., and Cai, H.N. (2003). The functional analysis of insulator interactions in the Drosophila embryo. Proc Natl Acad Sci U S A *100*, 5223-5228.

Nabirochkin, S., Ossokina, M., and Heidmann, T. (1998). A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. J Biol Chem *273*, 2473-2479.

Ong, C., Corces, V.G. (2009). Insulators as mediators of intra- and inter-chromosomal interactions: a common evolutionary theme. Journal of Biology *8*.

Pai, C.Y., Lei, E.P., Ghosh, D., and Corces, V.G. (2004). The centrolsomal protein CP190 is a component of the gypsy chromatin insulator. Molecular Cell *16*, 737-748.

Postow, L., Crisona, N.J., Peter, B.J., Hardy, C.D., and Cozzarelli, N.R. (2001). Topological challenges to DNA replication: conformations at the fork. Proc Natl Acad Sci U S A *98*, 8219-8226.

Ramos, E., Torre, E.A., Bushey, A.M., Gurudatta, B.V., and Corces, V.G. (2011). DNA topoisomerase II modulates insulator function in Drosophila. PLoS One *6*, e16562.

Roca, J. (1995). The Mechanisms of DNA Topoisomerases. Trends in Biochemical Sciences *20*, 156-160.

Roca, J., Berger, J. M. and Wang, J. C. (1993). J Biol Chem *268*, 5.

Rose, D., Thomas, W., and Holm, C. (1990). Segregation of recombined chromosomes in meiosis I requires DNA topoisomerase II. Cell *60*, 1009-1017.

Valenzuela, L., and Kamakaka, R.T. (2006). Chromatin insulators. Annu Rev Genet *40*, 107-138.

Wang, J.C. (1971). Interaction between DNA and an Escherichia coli protein. J Mol Biol *55*, 10.

Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. Nat Rev Mol Cell Biol *3*, 430-440.

Weil, R.a.V., J. (1963). The cyclis helix and cyclic coil forms of polyoma viral DNA. Proc Natl Acad Sci USA *50*, 8.