Published in final edited form as:

Chromosome Res. 2016 May ; 24(2): 175–181. doi:10.1007/s10577-015-9509-1.

ChIP bias as a function of cross-linking time

Laura Baranello^{1,*}, Fedor Kouzine¹, Suzanne Sanford¹, and David Levens¹

Laura Baranello: baranellof@mail.nih.gov; Fedor Kouzine: kouzinef@mail.nih.gov; Suzanne Sanford: ssanford926@gmail.com; David Levens: levens@helix.nih.gov

¹Laboratory of Pathology, NCI/NIH, Bethesda, MD, 20892, USA

Abstract

The chromatin immunoprecipitation (ChIP) assay is widely used to capture interactions between chromatin and regulatory proteins in vivo. Formaldehyde cross-linking of DNA and proteins is a critical step required to trap their interactions inside the cells before immunoprecipitation and analysis. Yet insufficient attention has been given to variables that might give rise to artifacts in this procedure, such as the duration of cross-linking. We analyzed the dependence of the ChIP signal on the duration of formaldehyde cross-linking time for two proteins: Topoisomerase 1 (Top1) that is functionally associated with the double helix in vivo, especially with active chromatin, and green fluorescent protein (GFP), that has no known bona fide interactions with DNA. With short time of formaldehyde fixation, only Top1 immunoprecipitation efficiently recovered DNA from active promoters, whereas prolonged fixation augmented non-specific recovery of GFP dramatizing the need to optimize ChIP protocols to minimize the time of cross-linking, especially for abundant nuclear proteins. Thus ChIP is a powerful approach to study the localization of protein on the genome when care is taken to manage potential artifacts.

Introduction

Chromatin immunoprecipitation is a valuable method for studying interactions between proteins and a genomic DNA region. As such it has provided fundamental insight into where and how gene regulatory processes occur in cells. The majority of ChIP applications use formaldehyde to cross-link reversibly protein to DNA or protein to protein [1, 2]. The cross-linked chromatin is fragmented, and protein-DNA complexes are then recovered by immunoprecipitation using an antibody that detects a chromatin-associated protein of interest. DNA sequences in the immunoprecipitate are then amplified by polymerase chain reaction (PCR). A basic premise of ChIP is that the cross-linking between double-strand DNA and proteins requires thermal base-flipping to expose the exocyclic amino groups on the bases that are otherwise hydrogen bonded [3]. Therefore the time of formaldehyde cross-linking is a critical parameter to determine the signal-to-noise ratio. Prolonged fixation would augment the non-specific recovery of soluble proteins not expected to associate with DNA, increasing the rate of false positives (i.e. protein-protein cross-linking).

Recent re-analysis of multiple results generated by ChIP-based assays has shown that control datasets (i.e., no immunoprecipitation and mock immunoprecipitation samples) often

* to whom correspondence should be addressed: Laura Baranello: baranellof@mail.nih.gov, Phone: 3014963552.

display patterns of recovery similar to “successful” experiments [4]. This issue could result from non-specific effect of DNA accessibility in different regions of the genome [5, 6]. Lastly, a completely heterologous protein in cells, GFP, could be preferentially enriched by a ChIP-based assay at highly active loci in an expression-dependent manner [7]. We noted that in these works, authors have used a rather long formaldehyde fixation time, ranging from 20 to 60 minutes, that might result in non-specific capturing of reactive soluble proteins at the open chromatin structure of active loci [5, 7]. To test whether the enrichment of non-specific protein factors is due to non-optimal ChIP conditions, we interrogated the binding of Top1 versus GFP as a function of cross-linking time.

A vector expressing either GFP fused to Top1 or GFP fused to a nuclear localization signal (NLS) was transfected into human HCT116 cells and after crosslinking for different time points ChIP'ed using GFP antibody. Quantitative PCR (qPCR) analysis of specific genes evidenced that non-specific recovery of GFP compared to Top1 is correlated with the duration of formaldehyde treatment.

Materials and Methods

Cell Cultures

Human colon cancer cells HCT116 were grown in DMEM supplemented with 10% heat-inactivated FCS.

Plasmid preparation and transient transfection

pMC-EGFP-Top1 was kindly donated by Dr. Prof Fritz Boege, Heinrich Heine University, Duesseldorf. pMC-EGFP-NLS was derived from the former by MluI and EcoRI digestion and ligation to the oligo - cg cgg ccg aaa aag aag cgt aaa gta taa cta gtt aa- which has a unique SpeI site. For chromatin immunoprecipitation, cells were seeded in 100 mm dish at a density of 70,000 cells per cm² 24 h before transfection. For each sample 1 to 12 µg of each plasmid were mixed with 30 µL of Lipofectamine 2000 in Optimem media according to the manufacturer's instructions. In parallel, cells were plated in 12 wells plate at the same density and transfected with the plasmids scaling the amount of material according to surface area. GFP expression was assessed by flow cytometry to define the amount of plasmid required to support equal GFP expression.

Flow cytometry

Samples were analyzed by flow cytometry to detect GFP signal on Cyflow ML Instrument (PARTEC) using FloMAX and/or FlowJo (Treestar version 7.6.1) software. At least 50,000 events were acquired for each sample.

Chromatin Immunoprecipitation (ChIP)

ChIP samples were prepared from HCT116 cells following Barski *et al.* protocol [8] with minor changes. Briefly, cells were cross-linked with 1% formaldehyde for 4, 10 or 60 minutes at 37°C or 25°C. Cross-linking was quenched by addition of 125 mM glycine. After harvesting cells, the pellet was resuspended in TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). Samples were sonicated and adjusted to the conditions of RIPA buffer by adding

1% Triton X100, 0.1% Na-Deoxycholate, 0.1% SDS and 200 mM NaCl. 7.5 µg of anti-GFP (ab290) or anti-Top1 (ab109374) or anti-RNA polymerase II (RNAPII) (ab817) or non-immune IgG (sc-2027, sc-2025) were mixed with Dynabeads Protein A and incubated at 4°C for 6 hours with rotation. Chromatin was added to the Protein A-antibody complexes and incubated overnight at 4°C with rotation. Immunoprecipitates were washed twice with RIPA buffer; twice with RIPA buffer plus 300 mM NaCl; twice with LiCl buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 250 mM LiCl, 0.5% NP40, 0.5% Na-Deoxycholate); and twice with TE. The beads were then resuspended in TE plus 0.25% SDS supplemented with proteinase K and incubated overnight at 65°C. After pooling, the DNA was purified with a Qiagen PCR extraction kit (QIAGEN). For qPCR detection performed using the LightCycler 480, the percent of IP enrichment as compared to input was calculated using FastStart DNA Master SYBR Green I kit (Roche Diagnostics). Quantification and melting curve analyses were performed using the Roche LightCycler software by the crossing point method as indicated by the supplier. All detection primers are listed in Table 1.

Results

Top1 is a nuclear protein that interacts extensively with DNA to release transcription-generated torsional stress [9, 10] and has been shown to bind transcribed promoters [11, 12]. The heterologous GFP on the contrary lacks any DNA binding domain and does not associate with the double helix. To minimize variations due to different antibody affinities we fused GFP to the Top1 gene or to a NLS. Transfected cells were cross-linked with formaldehyde for 4, 10 and 60 minutes prior to immunoprecipitation with an anti-GFP (or IgG) antibody and the recovery of GFP-Top1 at transcribed units was compared to the recovery of GFP-NLS. The crosslinking temperature likely affects the efficiency of formaldehyde reaction. However, to catch the in-vivo binding dynamics this temperature should be held close to the temperature of the growth medium. Consequently, some of the published ChIP protocols for mammalian cells recommend to use 37°C during incubation with the fixative, while others recommend to add formaldehyde directly to the cell media at 37°C and then incubate cells at room temperature (<http://www.emdmillipore.com>, Chromatin Immunoprecipitation protocol) [13]. The latter means that during initial stage of fixation the temperature of reaction is close to 37°C and then decrease. To avoid uncertainty with the changes of temperature during cross-linking, we fixed the cells by adding formaldehyde directly to the growth medium to a final concentration of 1% and incubated at 37°C. To confirm our conclusions, we repeated this experiment at condition when the temperature of growth medium and formaldehyde reaction were kept at 25°C.

To verify whether the level of DNA recovery depended on the accessibility of the chromatin, we used publicly available RNAPII ChIP-seq datasets (GSM935426) to identify a range of differentially expressed promoters (high and low RNAPII occupancy) (Figure 1A). We first confirmed (by RNAPII ChIP-qPCR) that the levels of RNAPII on these promoters were in accordance to the published dataset (Figure 1B). Then we proceeded to assess the level of GFP in the differentially transfected cells. Efficient recovery of DNA cross-linked with the transfected protein was not seen with GFP alone when crosslinking was performed for short time, but required fusion with Top1 (Figure 2A). The low level signal detected with GFP-NLS was not related to promoter strength (Figure 2A). However prolonged fixation with

formaldehyde (60 minutes) did not allowed to distinguish between specific and non-specific signal for low expressed promoters because it augmented the non-specific recovery of GFP (Figure 2B). Non-specific recovery of GFP was also observed at lower temperature (25°C) of cells fixation. qPCR analysis showed that the overall recovery was lower compared to crosslinking at 37°C but the GFP-Top1 vs GFP-NLS ratio was comparably high (Figure 3A). To rule out the possibility that a long exposure to formaldehyde could affect the structure of chromatin and in turn the occupancy of DNA-bound GFP-Top1, we ChIP'ed chromatin with anti-Top1. qPCR quantification revealed that longer fixation slightly increased the occupancy of Top1 at some promoters (independently from the level of expression) but overall did not affect the binding patterns of this enzyme (Figure 3B), as well as of another transcription factor such as RNAPII (Figure 1B).

Increasing the nuclear concentration of GFP should aggregate unspecific recovery of DNA fragments after immunoprecipitation. Predictably, non-specific over specific enrichment was increased together with nuclear concentration of GFP-Top1 and GFP-NLS after transfection of larger amount of plasmids (Figure 3C).

Discussion

Highly expressed genes are characterized by open chromatin conformation, nucleosome perturbations and DNA more exposed to interactions [14]. Under some cross-linking conditions this might cause the hyper-ChIP-ability of the regions [7]. However, because proteins accumulate at the promoters of expressed genes, eventually soluble factors such as GFP could be non-specifically trapped by over-cross-linking to other proteins (Figure 4); formaldehyde is one of the preferred histologic fixatives for tissue samples precisely because of its ability to cross-link with virtually all cellular proteins. One possible solution to the problem is to develop alternative approaches that avoid the cross-linking step [15, 16]. However, this lack of cross-linking necessarily means that only proteins very tightly associated with the chromatin can be immunoprecipitated. Another effective way to ascertain whether protein-DNA enrichment is biologically relevant is to fuse the protein of interest to GFP and compare the recovery of the fused protein to the one of GFP, as we did in this work.

Our data also demonstrate that non-specific cross-linking in human cell line is not an issue if formaldehyde treatment is limited to 10 minutes or less. Shortening the fixation time reduces the background level due to non-specific protein binding to trap preferentially native DNA-proteins interactions. This is in accordance with the results of Poorey *et al.* who demonstrated that bona fide *in vivo* cross-linking of transcription factors with DNA is rapid [17]. In addition, in some cases this could also prevent DNA damage response initiated inside the cells [18]. Consequently, some recently published results could be explained by prolonged formaldehyde treatment together with the small size of yeast cells [5, 7]. The correlation between non-specific recovery and duration of fixation together with nuclear concentration of protein observed in this study indicates that under some circumstances cross-linking artifacts might be relevant, but that good biochemical practice and proper controls can detect and minimize this bias.

Acknowledgments

Our research is supported by the Intramural Research Program of the US National Institutes of Health, Center for Cancer Research of the National Cancer Institute.

Abbreviations

ChIP	chromatin immunoprecipitation
GFP	green fluorescent protein
NLS	nuclear localization signal
PCR	polymerase chain reaction
qPCR	quantitative PCR
Top1	DNA topoisomerase 1
RNAPII	RNA polymerase II

References

1. Solomon MJ, Larsen PL, Varshavsky A. Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell*. 1988; 53(6): 937–47. [PubMed: 2454748]
2. Gilmour DS, et al. Topoisomerase I interacts with transcribed regions in *Drosophila* cells. *Cell*. 1986; 44(3):401–7. [PubMed: 3002635]
3. Gueron M, Kochoyan M, Leroy JL. A single mode of DNA base-pair opening drives imino proton exchange. *Nature*. 1987; 328(6125):89–92. [PubMed: 3037381]
4. Marinov GK, et al. Large-scale quality analysis of published ChIP-seq data. *G3 (Bethesda)*. 2014; 4(2):209–23. [PubMed: 24347632]
5. Fan X, Struhl K. Where does mediator bind in vivo? *PLoS One*. 2009; 4(4):e5029. [PubMed: 19343176]
6. Park D, et al. Widespread misinterpretable ChIP-seq bias in yeast. *PLoS One*. 2013; 8(12):e83506. [PubMed: 24349523]
7. Teytelman L, et al. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A*. 2013; 110(46):18602–7. [PubMed: 24173036]
8. Barski A, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007; 129(4):823–37. [PubMed: 17512414]
9. Baranello L, Kouzine F, Levens D. DNA topoisomerases beyond the standard role. *Transcription*. 2013; 4(5):232–7. [PubMed: 24135702]
10. Champoux JJ. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem*. 2001; 70:369–413. [PubMed: 11395412]
11. Puc J, et al. Ligand-dependent enhancer activation regulated by topoisomerase-I activity. *Cell*. 2015; 160(3):367–80. [PubMed: 25619691]
12. Kouzine F, et al. Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat Struct Mol Biol*. 2013; 20(3):396–403. [PubMed: 23416947]
13. Milne TA, Zhao K, Hess JL. Chromatin immunoprecipitation (ChIP) for analysis of histone modifications and chromatin-associated proteins. *Methods Mol Biol*. 2009; 538:409–23. [PubMed: 19277579]
14. Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin*. 2014; 7(1):33. [PubMed: 25473421]
15. Baranello L, et al. DNA break mapping reveals topoisomerase II activity genome-wide. *Int J Mol Sci*. 2014; 15(7):13111–22. [PubMed: 25056547]

16. O'Neill LP, Turner BM. Immunoprecipitation of native chromatin: NChIP. *Methods*. 2003; 31(1): 76–82. [PubMed: 12893176]
17. Poorey K, et al. Measuring chromatin interaction dynamics on the second time scale at single-copy genes. *Science*. 2013; 342(6156):369–72. [PubMed: 24091704]
18. Beneke S, et al. Chromatin composition is changed by poly(ADP-ribosyl)ation during chromatin immunoprecipitation. *PLoS One*. 2012; 7(3):e32914. [PubMed: 22479348]

Figure 1.

A) RNAPII occupancy at promoters and gene bodies of selected genes. Data are visualized as custom tracks in Genome Browser. **B)** RNAPII ChIP was performed and qPCR quantification at the same promoters as in **A)** confirmed the accordance between the two experiments. Data are shown as percentage of input DNA (n=2, error bars= s.d.). Mean recovery for non-immune IgG is 0.001.

Figure 2.

A) HCT116 cells were transfected with vectors expressing GFP-NLS or GFP-Top1 or mock. 24 hr following the transfection, cultures were analyzed by flow cytometry to verify that GFP-NLS and GFP-Top1 were expressed at the same level. Formaldehyde crosslinking was performed for 4, 10 and 60 minutes. Sonicated chromatin was immunoprecipitated with anti-GFP antibody or non-immune IgG. Recovered DNA was analyzed by qPCR on selected promoters (as in Figure 1). Data are shown as percentage of input DNA (n=3, error bars=s.d.). NT: non transfected control. * Low expressed gene. Mean recovery for non-immune IgG is 0.001. **B)** Comparison between different times of crosslinking. DNA recovery in the GFP-Top1 transfected cells was normalized for the DNA recovery in the GFP-NLS transfected cells (fold change) to better appreciate the specific promoter enrichment due to fusion with Top1.

Figure 3.

A) As in Figure 2A but formaldehyde crosslinking occurred at 25°C for 60 minutes. Data are presented as percentage of input DNA (n=3, error bars= s.d.). Mean recovery for non-immune IgG is 0.0008. **B)** ChIP enrichment of Top1 at selected promoter regions. Data are represented as percentage of DNA input (n=2, error bars=s.d.). **C)** Cells were transfected with GFP-NLS or GFP-Top1 (1X) and with 4 times increased amount of these vectors (4X). Cross-linking occurred for 10 minutes. qPCR at expressed and not expressed promoters shows the ratio of non-specific over specific enrichment.

Figure 4.
Non-specific chromatin binding of soluble proteins increases with long cross-linking time.

Table 1

List of qPCR primers.

Name	Forward Primer 5'-3'	Revers Primer 5'-3'
H3F3B	GGTAAGTTTCCCCGTCTGC	GCTCCTCACCTCCATTTCTG
HIF1a	AGCTCCTCAGTGACAGTGC	AGACTAGAGAGAAGCGGGCG
IRF1	AAGAGGGAAGAAGGCAGAGG	CTCGCCACTCCTTAGTCGAG
LOXL4	TCCACTTGAGCAAGAGCAAA	TGGCAGAGCAAGAGACAGAA
SKAP1	CATCACACATGGCAAGGAAC	GTCCACACAACCACATCTGC
CTAG2	CCACCTCTTGACCCTGTTGT	CTTCTGCGCAGGATGGAAG