Published in final edited form as:

Wiley Interdiscip Rev Syst Biol Med. 2009 November 1; 1(3): 400–406. doi:10.1002/wsbm.36.

Formaldehyde-Assisted Isolation of Regulatory Elements

Peter L. Nagy, MD, Ph.D.

peter-nagy@uiowa.edu; University of Iowa

David H. Price, Ph.D.

david-price@uiowa.edu; University of Iowa

Abstract

Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) is based on locus-specific variations in the ability of protein components of chromatin to trap genomic DNA following formaldehyde treatment. This variation is mostly due to uneven nucleosome distribution since histones are the most abundant and highly crosslinkable components of chromatin. The method can identify and enrich for physically accessible DNA segments of the eukaryotic genome corresponding to known regulatory regions and regions that might have thus far unidentified structural role in the nuclear organization of chromatin. The enrichment patterns are cell type specific and thus might provide information about how transcriptional systems are organized and regulated in various tissues and how they might be disrupted in disease states. Analysis of a 268 kb region of chromosome 19 in human fibroblasts shown here demonstrates that while most DNA fragments detected by FAIRE correspond to sites of DNaseI hypersensitivity in active regions of chromatin, some are found in otherwise repressed chromatin domains and at other sites that are not found with other methods used to probe chromatin structure. Further exploration of FAIRE is warranted due to the simplicity of the protocol and recent advancements in massively parallel sequencing.

Keywords

FAIRE; formaldehyde crosslinking; genome organization; chromatin structure; nucleosome distribution; regulatory sequences; transcription; DNase I sensitivity

Regulation of chromatin structure is an essential component of transcriptional control in eukaryotic cells [1]. Methods designed to study chromatin structure include Chromatin ImmunoPrecipitation (ChIP) [2], Chromosome Conformation Capture (3C) [3], DNase I and Micrococcal Nuclease (MNase) sensitivity studies [4-6], and a newly developed method called Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) [7,8]. What is known about FAIRE and how it complements the other methodologies used to study chromatin structure is summarized in this review. Every method has its strengths and weaknesses. ChIP involves formaldehyde crosslinking of living cells followed by fragmentation of their DNA and identification of DNA fragments associated with specific proteins enriched by affinity purification [9]. It allows precise mapping of the position of many chromosome components along DNA, although some factors seem resistant to formaldehyde crosslinking and cannot be localized by this method. ChIP also fails to provide information about the three dimensional arrangements of chromatin. 3C has been developed for the analysis of higher order chromatin [3]. In this method chromatin fixed by formaldehyde is restriction endonuclease treated and the ends generated are ligated together. Regions of the genome that are far apart along the DNA sequence but are juxtaposed due to looping of chromatin become contiguous and can be identified. ChIP and 3C are invaluable tools, but they do not answer what is the combined effect of various protein associations, specific modifications and three-dimensional arrangements on DNA accessibility.

Accessibility to soluble nuclear factors regulating transcription, repair and recombination is clearly of utmost importance. DNase I and MNase sensitivity studies provide such information, but they destroy accessible DNA fragments, and thus do not allow their isolation for further characterization. FAIRE is a method designed to identify and isolate specific genomic DNA sequences that are not readily trapped by formaldehyde crosslinking of chromatin [7,8]. Understanding the role of such genomic regions should provide insight into the organizational principles of chromatin.

The discovery of FAIRE

FAIRE is based on the fact that all regions of chromosomal DNA do not crosslink to chromosomal proteins equally well with formaldehyde. DNA segments that are trapped by crosslinked DNA binding proteins are retained in the interphase during phenol-chloroform extraction, while those DNA segments that are not protein associated accumulate in the aqueous phase. The method involves the following steps: 1) Formaldehyde crosslinking of the cells of interest. 2) Sonication to obtain DNA fragments a few hundred nucleotides long. 3) Phenol-chloroform extraction of the crosslinked sonicated material. 4) Precipitation of DNA enriched in the aqueous phase. 5) Identification of the DNA by microarray analysis or direct sequencing. The observation, that DNA fragments that crosslink poorly to proteins accumulate in the aqueous phase while the majority of DNA trapped by crosslinked protein components of chromatin forms a thick interphase, is hardly surprising [10]. To avoid loss of immunoprecipitated DNA to the phenol-chloroform interphase, ChIP protocols normally include overnight reversal of crosslinks before the immunoprecipitated material is phenol-chloroform extracted [2]. However, at the time of the discovery of FAIRE, it was not widely appreciated that DNA extracted from crosslinked chromatin would be qualitatively different from that obtained from non-crosslinked samples [7]. The original discovery of FAIRE was fortuitous and came during a ChIP-Chip (chromatin immunoprecipitation coupled with analyses of the enriched DNA fragments using genomic microarray) experiment to map the distribution of mono- di- and trimethylated histone tails in various mutants of the *S. cerevisiae* Set1 methyltransferase complex. Instead of using DNA extracted from untreated cells as a control, total DNA extracted from crosslinked cells was used as a reference for the ChIP'ed material. The result was a striking apparent enrichment for coding over non-coding regions in the immunoprecipitated material. Initially, this observation suggested that methylated nucleosomes were enriched in coding regions of the genome, however, similar results were obtained from mutant yeast strains that lacked H3K4 methylation. To obtain an explanation for this methylation independent enrichment of coding regions, material from every step was meticulously tested. It was concluded that the reference DNA isolated from crosslinked lysates was enriched for noncoding regions due to loss of coding regions trapped by crosslinked protein to the interphase during the phenol-chloroform extraction. Thus upon comparison of this reference with the ChIP-ed material, the latter appeared to be enriched for the coding regions. Later work showed that FAIRE was also able to detect variations in the crosslinkability of the significantly more complex human chromatin [8].

What does FAIRE detect?

The finding that noncoding regions of the yeast genome could be enriched by phenol/chloroform extraction of crosslinked and sonicated chromatin indicated that the crosslinkable protein concentration in these regions was lower than on the coding portion of RNA polymerase II transcribed genes. Formaldehyde penetrates organic materials quickly and forms stable but reversible methylene bridges, mainly between proteins, via the ϵ -nitrogen atom of lysine and an adjacent amide nitrogen of a peptide linkage. For DNA to react with formaldehyde, it must be partially denatured to expose the $-\text{CO-NH}$ grouping at the N-1 position of guanine, or the exocyclic amino groups of adenine, guanine, or cytosine

[11]. Because the bulk of chromatin is accounted for by the high lysine content and thus highly crosslinkable histone subunits of nucleosomes, it was proposed that the phenomenon that FAIRE detected uneven nucleosome distribution along chromatin [7]. Several other publications have since demonstrated that consistent with this hypothesis, regulatory regions in general have fewer nucleosomes than coding regions [12-15]. Enrichment for non-coding regions was more pronounced for highly transcribed genes, but was also obtained for silent genes [7]. Because the enrichment did not correlate with the level of transcription, the possibility of a role of specific DNA sequences or general base composition was examined [7,16-18]. However, no such correlation was found. A recent finding by Shivaswamy *et al.* provides an alternative explanation. These authors showed that specific nucleosomal rearrangements in promoters changed even in the absence of transcription [19]. They concluded that the relationship between chromatin remodeling and transcription is gene specific and that remodeling does not necessarily result in transcription. There are other data from the Lieb laboratory indicating that nucleosome dynamics in promoter regions is fundamentally different from that seen in coding regions [15]. In synchronized yeast cell populations progressing through the cell cycle, variations in crosslinkability are limited to promoters of cell cycle regulated genes.

Transcription factor occupancy also affects nucleosome density in regulatory regions. Destabilization of nucleosomes in the promoter regions of actively transcribed genes is well documented [5,6]. The question arises: why would it make a difference whether DNA is bound to histones or transcription factors? Shouldn't they be able to crosslink equally well? The answer to this question lies in the crosslinking characteristics of formaldehyde. Proteins that do not expose the exocyclic amino-groups of DNA do not crosslink to DNA directly, but depending on their steric relationship with DNA might simply trap it following intra-protein crosslinking [11]. In the original paper in which Solomon and Varshavsky proposed formaldehyde mediated DNA-protein crosslinking as a probe for *in vivo* analysis of chromatin structure, they show that purified *lac* repressor does not crosslink to *lac* operator-containing DNA [9]. They also obtained similar results with (A + T) DNA-binding protein (alpha-protein) to its cognate DNA.

It is also important to note that nucleosome density might not be the only determinant of chromatin crosslinkability. Methylation of the lysine residue K36 in the N-terminal tail of histone H3 (H3K36) has been shown to correlate extremely well with regulatory versus non-regulatory yeast chromatin [20]. This modification is most common in the body of transcribed genes and is responsible for recruitment of a histone deacetylase complex necessary for reestablishment of chromatin structure following passage of the polymerase [21]. Whether there is a causative relationship between FAIRE enrichment and the absence of the H3K36 dimethylation requires further study. Although methylation might affect the crosslinkability of specific lysine residues, it might have a more significant indirect effect through alteration of nucleosome compaction. FAIRE might be detecting the three-dimensional proximity of one nucleosome to others, irrespective of the linear distance between them. However, the degree of compaction of yeast chromatin is still unclear [18]. In higher eukaryotes the existence of histone variants that incorporate into chromatin in a replication independent manner could also potentially alter the ability of formaldehyde treated chromatin to trap associated DNA segments [22].

Functional characteristics of FAIRE enriched DNA sequences

FAIRE has been applied to analysis of both yeast and human chromatin. As discussed above, the original FAIRE studies in yeast demonstrated that there is an inherent difference between the crosslinkability of chromatin in coding regions versus noncoding regions. Giresi *et al.* performed a FAIRE study in human fibroblasts across the ENCODE regions

[23] which cover about 1% of the human genome [8]. In many genomic regions FAIRE fragments overlapped with DNase I hypersensitive sites, RNA polymerase II transcription start sites, and histone modifications associated with active transcription [4,24-27].

Figure 1 shows a 268 kb region of human chromosome 19q13.42 displaying the results obtained with different techniques designed to probe chromatin structure in a variety of different cell types. This region contains 18 annotated genes and a large number of repetitive elements. ChIP-chip results are shown from the Ludwig Institute which map the locations of RNA polymerase II, histone modifications associated with active chromatin (H3 acetylation and H3K4 dimethylation) as well as repressive chromatin marks (SUZ12 and H3K27 trimethylation) [24]. Data from Duke/NHGRI that maps sites of DNase I hypersensitivity [4] and FAIRE data from the Lieb laboratory [8] are also displayed. As described earlier [8], there is strong correlation between the localization of the FAIRE peaks with regions of DNase I hypersensitivity, as well as indicators of active chromatin (Fig. 1, between 59,295,000 and 59,410,000). On the contrary, FAIRE peaks are significantly reduced in regions of inactive chromatin defined by the lack RNA polymerase II and the presence of repressive chromatin marks (Fig. 1, <59,295,000 and >59,410,000). Interestingly, DNase I hypersensitivity correlates with FAIRE in both active and inactive chromatin (indicated by stars). The region of transition between active and inactive chromatin regions (59,380,000 to 59,430,000) was expanded and is shown in the lower panel of Figure 1. At this resolution it is clear that FAIRE correlates with transcriptional start sites, as was found earlier [8]. This is seen in the region containing three promoters in close proximity to each other (one for MBOAT7 and two for TSEN34) and on the promoter of the highly active ribosomal protein gene RPS9. In addition, FAIRE signals are relatively high throughout the coding region as well as downstream of the 3' end of the RPS9 gene. These regions are exactly where RNA polymerase II maps, suggesting that some FAIRE signals are due to displacement of nucleosomes by RNA polymerase II. However, FAIRE signals, accompanied by DNase I hypersensitivity sites, are also found in inactive chromatin (indicated by stars) demonstrating that not all FAIRE sites are caused by RNA polymerase II transcription.

Some FAIRE sites are considered “orphan” sites because they do not coincide with other marks of chromatin structure. Giresi *et al.* noted that the 40% of FAIRE peaks that fell into that category and attributed this finding to “the difference in cell types used among the experiments being compared and the sparse state of current human genome annotations” [8]. The functional importance of FAIRE elements is supported by the approximately 2-fold decrease in the frequency of insertion-deletion mutations in such regions compared to non-conserved, non-coding genomic regions [28]. To find a function for “orphan” FAIRE elements would be potentially helped by studies examining the localization of prominent FAIRE enriched regions in the nucleus. It would also be of interest to examine how FAIRE is affected by sudden changes in the nuclear compaction, for example during the activation of quiescent lymphocytes by phytohemagglutinins. Similarly, it would be of interest to examine how FAIRE is affected by transformation of cervical epithelial cells by human papilloma virus induced changes in nuclear morphology.

Practical uses of FAIRE

FAIRE is a simple, reproducible method that provides information about genomic organization as it is present in the cells at the time of fixation. Other methods, such as DNase I hypersensitivity studies require more extensive preparatory steps that allow compensatory transcriptional changes to occur and also increase the chance of unintended protein and DNA modifications and potentially degradation. There is also a need to perform pilot experiments to determine the activity of individual batches of the enzyme. FAIRE is not limited by availability of antibodies generated against a specific epitope of interest, or

the inability to obtain a functionally intact tagged version of the target protein. This strength of FAIRE is also its major weakness; the functional significance of the enrichment observed may be difficult to explain, unless there is data from other ChIP or hypersensitivity studies. FAIRE allows for physical isolation and identification by direct sequencing of genomic regions that are otherwise identified by their absence, such as DNase I hypersensitive sites. This in theory would make it possible to identify regulatory regions in organisms for which no sequence information is available. A caveat is that one study found that Negative Regulatory Elements (NRE), such as insulator elements, were not significantly enriched by FAIRE [29]. Yaragatti et al. [30] were successful using a method reminiscent of FAIRE to isolate regulatory regions of the human genome. They formaldehyde crosslinked F9 embryonal carcinoma cells, permeabilized their nuclei, and then used HaeIII digestion to fragment the genomic DNA trapped in the nucleus. Nucleosome free regions that were accessible to digestion produced small HaeIII fragments able to diffuse out of the nucleus, while DNA in tightly packaged, transcriptionally inactive regions was not released. They noted that there was some background release of DNA fragments that were associated with promoters inactive in F9 cells, which they removed using FAIRE. They used the DNA from the aqueous phase to generate a library that had about 20 times higher regulatory domain content than non-enriched DNA. In transcriptional reporter assays, about 20% of the fragments obtained showed significant promoter activity compared to the 1% in non-selected DNA. They were successful in identifying both promoter and enhancer regions using this methodology [30]. Another successful application of FAIRE was to demonstrate nucleosome deposition onto Cytomegalovirus (CMV) genomes following entry of the viral DNA into the nucleus [31]. Nucleosome-depleted chromatin was enriched by FAIRE at specific times following infection. The ratio of GAPDH and viral DNA recovered from fixed and nonfixed cells was compared using quantitative PCR studies using seven primer pairs corresponding to various functionally relevant regions of the viral genome. The results showed that the region corresponding to the viral replication origin (oriLyt) remained free of nucleosomes even when the rest of the virus became extensively chromatinized. There is reason to expect that similar studies will be conducted in the characterization of other viruses in the future.

Conclusion

FAIRE provides a snapshot view of the tremendously complex organization of chromatin. Most FAIRE enriched elements represent regulatory regions of the genome but some do not have easily attributable functional roles as of today. This might change as our understanding of chromatin structure grows, and as more people use the method to characterize various cell types and transcriptional states. The appearance of deep sequencing methodologies will allow better quantification of the enrichment accomplished by FAIRE and thus will allow a semiquantitative measurement of the crosslinkability of genomic DNA packaged into chromatin. The improved sensitivity and higher resolution promised by these technologies will allow for the use of FAIRE in characterization of the alterations in chromatin structure that is a hallmark of a number of genetic disorders such as laminopathies as well as various malignancies [32,33].

Acknowledgments

We thank Arkady Khodursky for critical reading of the manuscript. P.L.N. is supported by grant from the American Heart Association (0655618Z) and the NIH (NS064253). D.H.P. is supported by NIH (GM35500).

References

- [1]. Wallrath LL, Nagy PL, Geyer PK. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2008; 647:1–2.
- [2]. Lieb JD, Liu X, Botstein D, Brown PO. Nat Genet. 2001; 28:327–34. [PubMed: 11455386]
- [3]. Splinter E, Grosveld F, de Laat W. Methods Enzymol. 2004; 375:493–507. [PubMed: 14870685]
- [4]. Crawford GE, et al. Nat Methods. 2006; 3:503–9. [PubMed: 16791207]
- [5]. Keene MA, Elgin SC. Cell. 1981; 27:57–64. [PubMed: 6799212]
- [6]. Keene MA, Corces V, Lowenhaupt K, Elgin SC. Proc Natl Acad Sci U S A. 1981; 78:143–6. [PubMed: 6264428]
- [7]. Nagy PL, Cleary ML, Brown PO, Lieb JD. Proc Natl Acad Sci U S A. 2003; 100:6364–9. [PubMed: 12750471]
- [8]. Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD. Genome Res. 2007; 17:877–85. [PubMed: 17179217]
- [9]. Solomon MJ, Varshavsky A. Proc Natl Acad Sci U S A. 1985; 82:6470–4. [PubMed: 2995966]
- [10]. Keeney S, Kleckner N. Proc Natl Acad Sci U S A. 1995; 92:11274–8. [PubMed: 7479978]
- [11]. Feldman MY. Prog Nucleic Acid Res Mol Biol. 1973; 13:1–49. [PubMed: 4573489]
- [12]. Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD. Nat Genet. 2004; 36:900–5. [PubMed: 15247917]
- [13]. Bernstein BE, Liu CL, Humphrey EL, Perlstein EO, Schreiber SL. Genome Biol. 2004; 5:R62. [PubMed: 15345046]
- [14]. Sekinger EA, Moqtaderi Z, Struhl K. Mol Cell. 2005; 18:735–48. [PubMed: 15949447]
- [15]. Hogan GJ, Lee CK, Lieb JD. PLoS Genet. 2006; 2:e158. [PubMed: 17002501]
- [16]. Dekker J. Genome Biol. 2007; 8:R116. [PubMed: 17577398]
- [17]. Gupta S, Dennis J, Thurman RE, Kingston R, Stamatoyannopoulos JA, Noble WS. PLoS Comput Biol. 2008; 4:e1000134. [PubMed: 18725940]
- [18]. Dekker J. J Biol Chem. 2008; 283:34532–40. [PubMed: 18930918]
- [19]. Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR. PLoS Biol. 2008; 6:e65. [PubMed: 18351804]
- [20]. Rao B, Shibata Y, Strahl BD, Lieb JD. Mol Cell Biol. 2005; 25:9447–59. [PubMed: 16227595]
- [21]. Carrozza MJ, et al. Cell. 2005; 123:581–92. [PubMed: 16286007]
- [22]. Henikoff S, Ahmad K. Annu Rev Cell Dev Biol. 2005; 21:133–53. [PubMed: 16212490]
- [23]. Birney E, et al. Nature. 2007; 447:799–816. [PubMed: 17571346]
- [24]. Kim TH, et al. Genome Res. 2005; 15:830–9. [PubMed: 15899964]
- [25]. Kim TH, et al. Nature. 2005; 436:876–80. [PubMed: 15988478]
- [26]. Sabo PJ, et al. Nat Methods. 2006; 3:511–8. [PubMed: 16791208]
- [27]. Koch CM, et al. Genome Res. 2007; 17:691–707. [PubMed: 17567990]
- [28]. Clark TG, Andrew T, Cooper GM, Margulies EH, Mullikin JC, Balding DJ. Genome Biol. 2007; 8:R180. [PubMed: 17784950]
- [29]. Petrykowska HM, Vockley CM, Elnitski L. Genome Res. 2008; 18:1238–46. [PubMed: 18436892]
- [30]. Yaragatti M, Basilico C, Dailey L. Genome Res. 2008; 18:930–8. [PubMed: 18441229]
- [31]. Nitzsche A, Paulus C, Nevels M. J Virol. 2008; 82:11167–80. [PubMed: 18786996]
- [32]. Filesi I, et al. Physiol Genomics. 2005; 23:150–8. [PubMed: 16046620]
- [33]. Zink D, Fischer AH, Nickerson JA. Nat Rev Cancer. 2004; 4:677–87. [PubMed: 15343274]

Figure 1. Comparison of FAIRE with other techniques probing chromatin structure

A region of human chromosome 19 was analyzed using the UCSC genome browser (<http://genome.ucsc.edu/>) as described in the text. Numbers at the top of both panels indicate chromosomal coordinates. In both panels the positions of genes in the regions covered, as well as the locations of repeated sequences are indicated. Stars denote position of FAIRE signals that overlap with sites of DNase I hypersensitivity in inactive chromatin.