ISSN 0006-3509, Biophysics, 2012, Vol. 57, No. 2, pp. 140–143. © Pleiades Publishing, Inc., 2012.
Original Russian Text © A.A. Belostotsky, 2012, published in Biofizika, 2012, Vol. 57, No. 2, pp. 215–220.

MOLECULAR BIOPHYSICS

Analysis of Protein-on-DNA Binding Profiles Determined with ChIP-seq Reveals Possible Interaction of Specific Transcription Factors with RNA Polymerase II during Transcription Elongation*

A. A. Belostotsky

Institute for Genetics and Selection of Industrial Microorganisms, Moscow, 117545 Russia

Received February 2, 2012

Abstract—It is considered that almost all transcription factors stay on promoter while RNA polymerase II "clears" the promoter and "proceeds" to elongation transcribing mRNA. However, analysis of some specific transcription factors and RNA polymerase II binding profiles on DNA, detected with ChIP-seq method, revealed possibility of interaction between transcription factors and RNA polymerase II in the process of transcription elongation.

Keywords: transcription initiation, transcription elongation, RNA polymerases, transcription regulators.

DOI: 10.1134/S0006350912020054

INTRODUCTION

The process of transcription involves RNA polymerase (Pol) and a multitude of regulatory proteins ensuring RNA synthesis in certain genome regions under certain conditions [1]. Among such proteins there are transcription initiation factors that bind to certain DNA sequences and to Pol directly or via other proteins. For Pol II, are transcription elongation factors [2] and splicing factors, binding to it and regulating the rate of respective processes [3]; and transcription termination factors, involved in termination of mRNA synthesis and transcript release [4] also are well known.

Despite the close structural and regulatory «interlacing» of transcription initiation and elongation [5, 6], it is thought that almost all initiation factors are associated with the promoter and do not move together with Pol II during elongation. Initiation and elongation are separated even by competition of some respective factors for a common binding site on Pol II [7]. Proteins commonly referred to as transcription factors without further specification are usually, by default, also taken to remain on the promoter or on distal regulatory elements, known as enhancers. In fact, they are treated as *initiation* factors of varying degree of regulatory specificity, activating or in some cases inhibiting the transcription of a certain set of genes. Here it is shown that transcription factors may be bound with Pol II in the course of elongation as well. This conclusion is drawn from analysis of the ChIP-seq signals.

The ChIP-seq technique is based on isolation and identification of DNA regions co-immunoprecipated with protein complexes [8]. As a result, any DNA region can be assigned a measure of its association with a complex in cell culture at the moment of experiment. The spectrum of values thus obtained has the obvious sense of a protein-on-DNA binding profile. It is conventionally regarded only as raw material to be processed for isolating large peaks. However, a more important aspect of using ChIP-seq to elucidate the structure of complexes, their binding and dynamics on DNA is analysis of the *off-peak* signal.

In a ChIP-seq experiment, proteins are cross-linked with DNA and also with each other. This circumstance is crucial, signifying that an antibody that is specific against one protein can precipitate not only the DNA region bound with this protein but also the DNA regions bound with proteins that are in complex with the antibody target protein. In this way the inverse problem can be solved, inferring that proteins with similar binding profiles belong to one complex.

Therewith the complex is assumed to be intact through the procedure, and binding of free proteins or complexes to DNA during isolation is negligible. Also the relative specificity of antibodies, the lack of conformational preference, and the absence of spatial shielding of proteins in complex are assumed. These are the common, albeit usually implicit, assumptions of ChIP-seq and immunoprecipitation in general.

* The experimental data contained herein fully correspond to the original publication but the text was substantially revised for the English version. *A.G.*

DATA AND METHODS

General data and tools. The initial data were the ChIP-seq signals for transcription factors and Pol II and the processed peaks thereof, available from the UCSC Genome Browser (http://genome.ucsc.edu/) [9, 10]. The Tables service (http://genome.ucsc.edu/cgi-bin/hgTables) [11] was used to determine all parameters (the number of points, the value range, the mean and the standard deviation), to define intersections and estimate correlations. Because of a bug in the service, the table of annotated genes (Knowngene) had to be exported from the database, the required columns txStart and txEnd together with the Chrom column were excised and imported back into the database as a Custom Track. This provided adequate search for intersections of the signal with the annotated genes, which represent the "intragenic" ChIP-seq signals. The exons from the Knowngene sample were treated in the same way.

Further, the peak data as such are not necessary for the probative part of this work and are not presented here; some of them are just mentioned.

Initial data specification. All transcription regulators for which ChIP-seq data had been obtained on HeLa in the Encode project (http://genome.ucsc.edu/ENCODE/) [12] were considered. These were proteins regarded as transcription factors for Pol II: AP-2a, AP-2g, c-Fos, c-Jun, c-Myc, E2F1, E2F4, E2F6, HA-E2F1, junD, Max, Nif1, TR4; subunits of the SWI/SNF chromatin remodeling complex: BAF155, BAF170, Ini1, Brg1; and transcription initiation factors for Pol III (used as control): BDP1, BRF1, BRF2, RPC155, TFIIIC. The initial files are available at http://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeYaleChIPseq/. The data on Pol II binding in GM12878 cell culture were also used.

The model DNA region considered here is the smallest human autosome, Chr22; genome assembly hg18. A 10 MB filter was used with very large datasets. Samples of 100 KB were also tested in order to assess the effects of sample volume.

RESULTS AND DISCUSSION

Pol II transcription regulators bind within genes (especially exons) more than beyond genes. It is supposed that the regulators associated with Pol II during elongation would thereby be "carried away" from the initial site of binding on DNA, and thus would be detected by ChIP-seq as those associated with intragenic regions. Note that this analysis concerns not the stationary binding (in peaks) but rather a dynamic association, "transitory" binding (off-peak).

Thus it can be expected that the mean off-peak signal of Pol II regulators within the genome territories actively transcribed by this enzyme (such as annotated genes in the present study) would exceed the mean signal outside such regions.

And indeed, for all Pol II transcription regulators but one (E2F4) the off-peak signal within genes was somewhat higher than outside genes (not shown). Yet the difference in signals was much smaller than could be expected. Even considering that the genome is not fully annotated and thus some genes are unwittingly entered into the "intergenic" subsample, a few percent difference falls short of a pronounced effect. A more marked difference in intragenic and intergenic signals was observed for *regulated* and *unregulated* genes (data not shown); the former were determined as intersections with the peaks of each regulator separately.

Considering that exons are transcribed more slowly than introns, there is reason to compare the exonic ChIP-seq signals with those beyond, i.e. in introns and "intergenic" regions (though the latter, as just admitted, may contain unidentified genes and their "coding status" is thus uncertain).

Just as implied by the idea of persistent association of transcription factors with the elongating Pol II, the mean exonic off-peak signal proved to be obviously higher than the non-exonic one (Fig. 1). The difference was much greater than that for whole genes versus intergenic regions, and the only artifact found in the latter comparison (the E2F4 case) was rectified. Thus, a fully consistent pattern has been obtained: the Pol II transcription regulators "transitorily" reside within the exons to a substantially greater extent than outside them.

The regulators of Pol III (right-hand part of Fig. 1) provide quite a proper control in this respect; the non-exonic off-peak signal of Pol II itself can also be regarded as background.

Further, it is obvious that in exons, all Pol II regulators except c-Fos are more abundant than those of Pol III, which is also quite a reasonable result. Hardly any such distinction could be observed with whole genes. This again points to the inadequacy of considering simply intragenic versus extragenic ChIP-seq signals, because for the former the calculated mean would be severely affected by extensive low-signal introns.

The relationship between mean exonic and non-exonic signals was of course largely retained in the case of regulated genes (data not shown). The difference was more pronounced for a larger sample.

Binding of Pol II and its transcription factors within genes (exons) is positively correlated. Neither the level of ChIP-seq signals nor the co-localization of peaks give proof of spatiotemporal association of DNA-binding proteins; they can only show coincidence in space. Nonetheless, one can search for "subtler" effects that would be indicative of some structural unity, holding different parts together both in space and in time. In the present case, there is reason to consider *preferential binding* of different regulators of

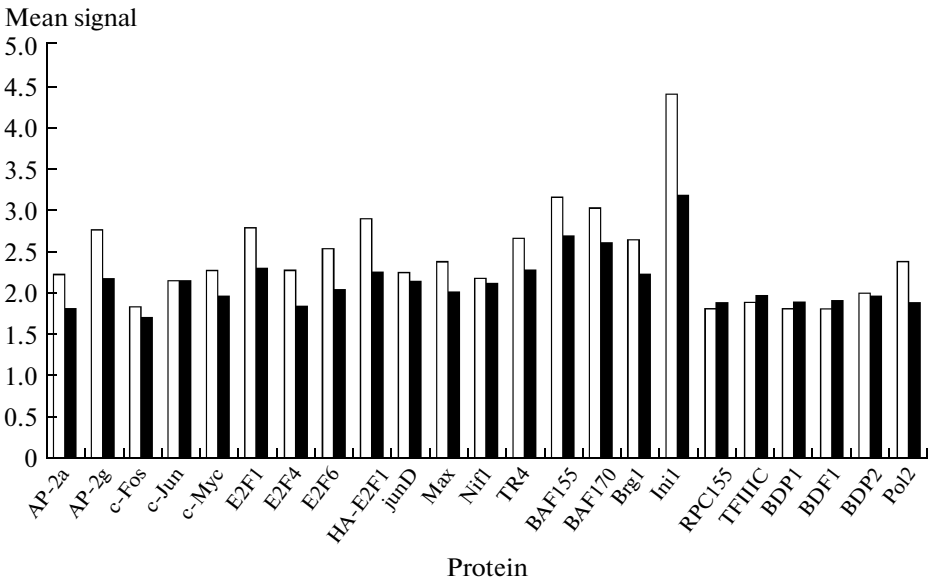

Fig. 1. Mean off-peak ChIP-seq signals of transcription regulators within exons (light bars) and outside exons (dark bars).

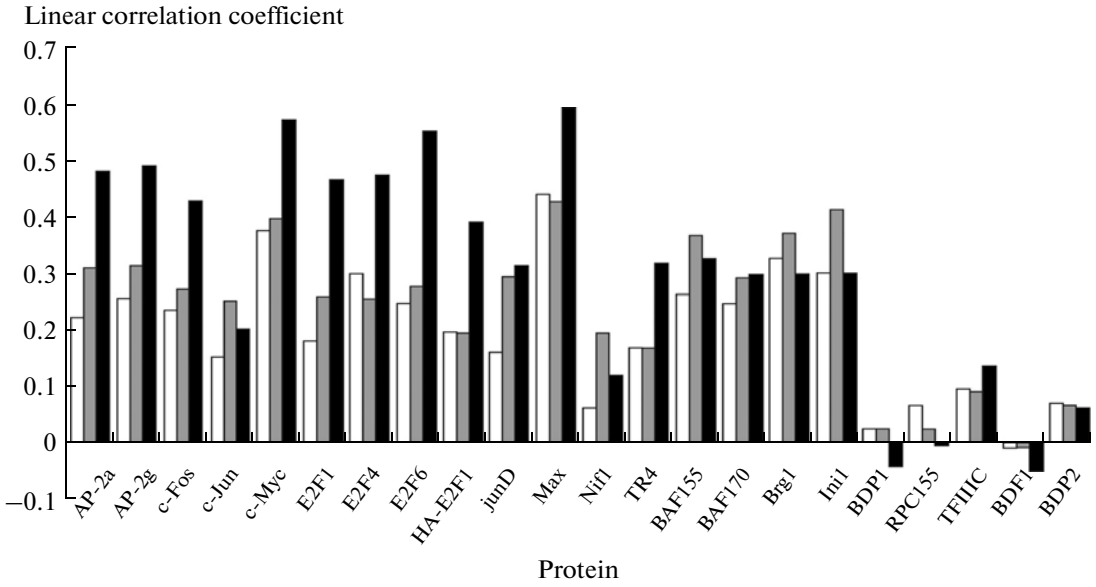

Fig. 2. Correlations of exonic ChIP-seq signals: (light bars) Pol II off-peak vs. regulator off-peak, (gray bars) Pol II off-peak vs. regulator whole signal, (black bars) Pol II whole signal vs. regulator off-peak.

Pol II *with the regions* of DNA within a gene (exon). Similarity of the corresponding binding profiles would be a strong enough argument in favor of structural association of transcription initiation factors with the elongating Pol II. Finally, it was expedient to examine the correlations of ChIP-seq signals.

Specifically, linear correlation coefficients (in the standard formulation) were determined for the binding profiles of transcription regulators and Pol II, using 1-nt scanning.

Quite similar results have been obtained with genes and with exons; the latter are shown in Fig. 2 for different combinations of signals. The pronounced positive correlations observed for most of the Pol II regulators (except Nif1 and perhaps c-Jun) strongly suggest their co-localization with the enzyme. Again, the Pol III transcription factors with small and sign-changing correlation coefficients make a negative control group.

On the whole, the correlation for Pol II and its transcription factors (left half of Fig. 2) is no less pro-

nounced than that for Pol II and the SWI/SNF subunits (BAF155, BAF170, Brg1, Ini1). In this analysis, the latter group can be regarded as a kind of positive control, inasmuch as the SWI/SNF remodeler is held not only to associate with Pol II on the promoter (which was confirmed here; data not shown) but also to attend Pol II in elongation [13]. Thus, there are grounds for supposing that a number of Pol II transcription factors do "co-migrate" with the enzyme along the DNA, which implies that some structural association between them (at least indirect) is maintained during elongation.

It is noteworthy that the correlation beyond a gene and within it turned out to be practically the same. Further, the correlation between regulator and Pol II signals in unregulated genes proved even somewhat higher than in regulated ones; the same was observed in respective exons. This appears to be a normal phenomenon, being also true for the "positive controls" BAF155 and BAF170. Intriguingly, the highest correlations in the Pol II transcription factor group were those between the enzyme signal in the whole exon and the regulator signal off-peak (black bars in Fig. 2). Here it should be borne in mind that weak signals can strongly correlate with intense ones owing to even small "common features" connected with a systematic error of the method. Finally, correlations within peaks that are often negative despite obvious co-localization of the peaks (extensive statistics is available) are just a property of calculating the linear correlation coefficient.

Notwithstanding, the main established fact is that the high signal of Pol II in genes and especially in exons is stably positively correlated exclusively with the signals of specific Pol II regulators.

Note also that such correlations may in a certain sense be cell(tissue)-specific: the correlation between a Pol II regulator signal in HeLa and the Pol II enzyme signal in GM12878 is substantially weaker than those within each cell line (not shown). This may be indirect evidence in favor of general specificity and biological relevance of the alleged association.

CONCLUSIONS

Analysis of the ChIP-seq signals provides data in support of the idea that most of the transcription factors engaged in the initiation of Pol II transcription also attend the elongation stage. The mean signals of Pol II and relevant transcription regulators within the genes regulated thereby or their exons prove to be higher than outside such genes (exons). The pronounced positive correlation between the Pol II and transcription factor signals outside the large binding peaks in genes suggests persistent structural association of these components throughout transcription. Note that the reasoning here involves no other assumptions than the overall adequacy of the ChIP-seq approach and the applicability of linear correlation analysis.

Conceptually, the conclusions made in the present work agree most nicely with the existence/operation of "stationary" transcription factories (see, e.g., [14]). Testing this hypothesis and building a holistic model also can and should be approached with analysis of the "dynamic" ChIP-seq profiles; this will be the subject of further work.

REFERENCES

1. R. G. Roeder, FEBS Lett. **579**, 909 (2005).
2. K. M. Arndt and C. M. Kane, Trends Genet. **19**, 543 (2003).
3. T. Misteli and D. L. Spector, Mol. Cell. **3**, 697 (1999).
4. P. Richard and J. L. Manley, Genes Dev. **23**, 1247 (2009).
5. D. M. Prather, E. Larschan, and F. Winston, Mol. Cell Biol. **25**, 2650 (2005).
6. B. Guglielmi, J. Soutourina, C. Esnault, and M. Werner, Proc. Natl. Acad. Sci. USA **104**, 16062 (2007).
7. D. Grohmann, J. Nagy, A. Chakraborty, et al., Mol. Cell. **43**, 263 (2011).
8. E. R. Mardis, Nat. Methods **4**, 613 (2007).
9. W. J. Kent, C. W. Sugnet, T. S. Furey, et al., Genome Res. **12**, 996 (2002).
10. P. A. Fujita, B. Rhead, A. S. Zweig, et al., Nucl. Acids Res. 2010, Oct 18 [Epub ahead of print].
11. D. Karolchik, A. S. Hinrichs, T. S. Furey, et al., Nucl. Acids Res. **32** (Database issue), 493 (2004).
12. K. R. Rosenbloom, T. R. Dreszer, M. Pheasant, et al., Nucl. Acids Res. **38** (Database issue), 620 (2010).
13. M. A. Schwabish and K. Struhl, Mol. Cell Biol. **27**, 6987 (2007).
14. J. A. Mitchell and P. Fraser, Genes Dev. **22**, 20 (2008).