Phys. Biol. **8** (2011) 044001 (4pp) doi:10.1088/1478-3975/8/4/044001

COMMUNICATION

# Nucleosome mediated crosstalk between transcription factors at eukaryotic enhancers

**Vladimir B Teif and Karsten Rippe**

BioQuant and German Cancer Research Center, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

E-mail: Vladimir.Teif@bioquant.uni-heidelberg.de and Karsten.Rippe@dkfz.de

Received 5 February 2011
Accepted for publication 24 May 2011
Published 10 June 2011
Online at stacks.iop.org/PhysBio/8/044001

**Abstract**

A recent study of transcription regulation in *Drosophila* embryonic development revealed a complex non-monotonic dependence of gene expression on the distance between binding sites of repressor and activator proteins at the corresponding enhancer *cis*-regulatory modules (Fakhouri *et al* 2010 *Mol. Syst. Biol.* **6** 341). The repressor efficiency was high at small separations, low around 30 bp, reached a maximum at 50–60 bp, and decreased at larger distances to the activator binding sites. Here, we propose a straightforward explanation for the distance dependence of repressor activity by considering the effect of the presence of a nucleosome. Using a method that considers partial unwrapping of nucleosomal DNA from the histone octamer core, we calculated the dependence of activator binding on the repressor–activator distance and found a quantitative agreement with the distance dependence reported for the *Drosophila* enhancer element. In addition, the proposed model offers explanations for other distance-dependent effects at eukaryotic enhancers.

S Online supplementary data available from stacks.iop.org/PhysBio/8/044001/mmedia

## Introduction

Predicting gene expression from the DNA sequence and arrangement of regulatory proteins on the DNA is a central issue of the current research in quantitative cell biology (Yuh *et al* 1998, Beer and Tavazoie 2004, Jaeger *et al* 2004, Janssens *et al* 2006, Zinzen *et al* 2006, Yuan *et al* 2007, Segal *et al* 2008, Gertz *et al* 2009, He *et al* 2010, Kaplan *et al* 2011). The underlying models are usually constructed assuming the competitive equilibrium binding of multiple proteins at genomic regulatory regions (Bintu *et al* 2005a, 2005b, Garcia *et al* 2010, Teif 2010). However, the main complication encountered in eukaryotes is the organization of the DNA genome in chromatin: about 145–147 bp of DNA are wrapped around a histone octamer protein core to form a nucleosome chain with 10–50 bp linker DNA spacing. Thus, DNA is not equally accessible for transcription factors as assumed in the early models. Integrating the chromatin structure within the framework of probabilistic transcription factor binding is still an unsolved problem that is highly relevant to rationalize the complexity and cooperativity of protein interactions in the genome. Current computational models usually derive a phenomenological potential for TF–TF interactions from fitting the experimental data (He *et al* 2010). However, the predictive power of such approaches is limited since they lack mechanistic molecular details of the underlying processes. Here we address a recent experimental study of *Drosophila* embryonic development, which considered synthetic enhancers with varying distance between binding sites for a repressors/activator transcription regulation module (figure 1(*A*)) (Fakhouri *et al* 2010). Intuitively, one might expect that the effect of the repressor would simply decrease with its target distance. However, the study by Fakhouri *et al* revealed a puzzling non-monotonic distance dependence of

 

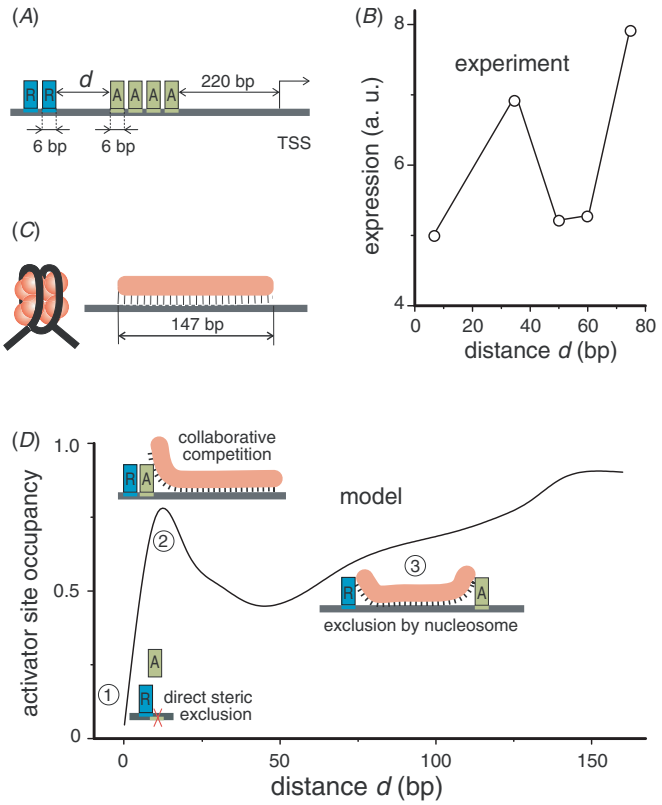


**Figure 1.** (*A*) Experimental setup investigated (Fakhouri *et al* 2010). The enhancer element contains two binding sites for repressor '*R*' and four for activator '$A_1$–$A_4$'. The distance *d* is varied. (*B*) Gene expression as a function of *d* plotted as $1 - P$, with *P* being the repressor quenching efficiency reported by Arnosti and co-workers (Fakhouri *et al* 2010). (*C*) A lattice model of the nucleosome that allows DNA unwrapping with 1 bp resolution (Teif *et al* 2010). (*D*) Activator binding to the leftmost '*A*' site calculated as a function of distance *d* assuming that the contact of nucleosome '*N*' with repressor is cooperative ($w = 1000$). Repressor and activator cover 6 bp upon binding to DNA and exclude each other at distances $< 6$ bp. Binding constants: $K(R) = 10^{11}$ M$^{-1}$; $K(A_1) = K(A_2) = 5 \times 10^9$ M$^{-1}$, $K(A_3) = K(A_4) = 10^{10}$ M$^{-1}$; $K(N) = 10^{-8}$ M$^{-1}$. Concentrations: $[R] = [A] = [N] = 10^{-9}$ M. Nucleosomes form nonspecifically, cover up to 147 bp and can partially unwrap with a homogeneous unwrapping potential (Teif *et al* 2010).

the repression efficiency. The repressor efficiency was high at small separations ∼5 bp, low around 30 bp, reached a maximum at 50–60 bp, and decreased at larger distances to the activator binding sites (figure 1(*B*)). How can the observed distance dependence between activator and repressor binding sites be rationalized? Here, a quantitative explanation of these findings is proposed that takes into account the nucleosomal chromatin structure.

## Results

A number of explanations of the distance dependence observed by Fakhouri *et al* can be excluded: (i) proteins with extended tails may interact in a distance-dependent manner but for such large distances that would be unprecedented since these interactions are usually ⩽15 bp (Teif 2007). (ii) A third 'mediator' protein (complex) could insert between the repressor and the activator to quench the activator. The transcription factors in question interact with multicomponent co-repressors and co-activators. Thus, a flexible assembly of proteins with possibly multiple contact surfaces might indeed bridge relatively large distances through direct touching. However, the wide range of distances over which the repressor acts suggests a general mechanism not dependent on a repressor complex with fixed geometry. (iii) Another source of nonlinearity could arise through protein-assisted DNA looping. However, the probability of loop formation has a maximum at 500 bp separation (Rippe 2001) and a 10 bp periodicity below the DNA persistence length of 150 bp (Saiz *et al* 2005). In contrast, the experimental data of Fakhouri *et al* show a peak at 50–60 bp separation, which would be energetically unfavorable for interactions via the looping of a relatively stiff free DNA tether. Thus, we are left with the possibility that the repressor acts indirectly through chromatin rearrangements. Such an assumption is indeed supported by a subsequent recent publication by Li and Arnosti (2011). It was found that upon adding short-range repressors enhancer regions become less susceptible to MNase digestion and histone deacetylation increases. While the latter would increase nucleosome stability (Teif and Rippe 2010), the relation of this process to gene expression remains enigmatic.

Here, we propose a straightforward mechanistic explanation for the distance dependence of repressor action by considering the nucleosome structure of the eukaryotic genome. Specifically, we assume that the repressor binding stabilizes the nucleosome. This can be realized either by the direct repressor interaction with the core nucleosomal DNA (Dowell *et al* 2010) or by recruiting histone modifying enzymes (Teif and Rippe 2010). Although the nucleosome is stabilized, its DNA can partially unwrap to allow activator binding. Using a novel method that considers partial unwrapping of nucleosomal DNA (Teif *et al* 2010) (figure 1(*C*)), we calculated the dependence of activator binding on the repressor–activator distance *d* (figure 1(*D*)). A fixed set of reasonable thermodynamic parameters detailed in the figure legend was applied without further fitting. It was assumed that nucleosomes can assemble at any position along the DNA and can partially unwrap as described by a homogeneous potential (Teif *et al* 2010). Repressor–nucleosome contacts were included via a McGhee–von Hippel cooperativity parameter (McGhee and von Hippel 1974) as described previously (Teif 2007). Furthermore, it was assumed that the simultaneous binding of repressor and activator was prohibited if binding sites were within less than 6 bp from each other. The details of the calculation method are provided in the supplementary materials available at stacks.iop.org/PhysBio/8/044001/mmedia.

Our calculations show that three distance-dependent regimes emerge for such a system, labeled as 1, 2 and 3 in figure 1(*D*). At small separations ($d < 6$ bp, regime 1), activator binding is inhibited directly by the repressor. At 6 bp $< d <$ 50 bp (regime 2), repressor and activator bind cooperatively. The binding of one protein stabilizes the unwrapped state of nucleosomal DNA and facilitates the binding of the second

protein in a so-called collaborative competition (Miller and Widom 2003). This effect decreases with the repressor–activator distance. At larger distances of 50 bp $< d <$ 147 bp (figure 1(*D*), regime 3), a partially unwrapped nucleosome can fit between the repressor and activator making activator binding anticooperative. This effect decreases at larger distances as less unwrapping of nucleosomal DNA is required. Thus, our calculations reproduce the essential features of the nonlinear distance dependence observed experimentally (Fakhouri *et al* 2010) (figures 1(*B*) and (*D*)). The maximum repressor efficiency at 50–60 bp distances corresponds to the inner DNA region in the central part of nucleosome, which is the least accessible to transcription factors. The distance corresponding to the maximum repressor efficiency is mainly determined by the nucleosome geometry and its possibility to unwrap (supplementary figure S2 available at stacks.iop.org/PhysBio/8/044001/mmedia). In contrast, the width and the height of the peak are sensitive to changes in the binding constants of repressor, activator and histone octamer (supplementary figure S3 available at stacks.iop.org/PhysBio/8/044001/mmedia).

## Discussion

Cooperative interactions between transcription factors separated by distances less than the nucleosome size seem to be abundant in the eukaryotic genome (Segal *et al* 2008). With respect to *Drosophila* enhancers, three classes of preferred distances between TF binding sites can be roughly distinguished: class I has ∼10 bp separations between homotypic TFs, probably reflecting their in-phase arrangement on the same side of the double helix (Makeev *et al* 2003); class II shows an ∼17 bp separation between heterotypic TFs that are probably located at the opposite sides of the double helix (Makeev *et al* 2003, Papatsenko *et al* 2009); class III is characterized by preferred distances between activators and repressors centered at about 60–80 bp (Papatsenko *et al* 2009). The latter distance cannot be rationalized in terms of the DNA double helix phasing and is likely reflecting structural chromatin features. Our calculations suggest that this class of distance preferences represents regulatory elements that operate via nucleosome-mediated TF interactions. For this class, we have provided here a quantitative description of the nucleosome-dependent regulation of gene expression at short genomic distances. The nucleosome in the lattice model is represented as a protein entity with a characteristic size of the histone octamer that can be wrapped by 147 or less DNA base pairs. Although histones are the most abundant chromatin proteins, one could also imagine protein complexes other than histones, which form an 'enhanceosome' with mathematical properties similar to the nucleosome in terms of DNA accessibility. The nonlinear distance dependence predicted by our model would be present in both cases.

## Acknowledgments

We thank David Arnosti and Thomas Höfer for fruitful discussions. This work was funded within project EpiGenSys by the BMBF as a partner of the ERASysBio+ initiative supported under the EU ERA-NET Plus scheme in FP7 and by BRFFI grant B10M-060.

## References

Beer M A and Tavazoie S 2004 Predicting gene expression from sequence *Cell* **117** 185–98

Bintu L, Buchler N E, Garcia H G, Gerland U, Hwa T, Kondev J, Kuhlman T and Phillips R 2005a Transcriptional regulation by the numbers: applications *Curr. Opin. Genet. Dev.* **15** 125–35

Bintu L, Buchler N E, Garcia H G, Gerland U, Hwa T, Kondev J and Phillips R 2005b Transcriptional regulation by the numbers: models *Curr. Opin. Genet. Dev.* **15** 116–24

Dowell N L, Sperling A S, Mason M J and Johnson R C 2010 Chromatin-dependent binding of the *S. cerevisiae* HMGB protein Nhp6A affects nucleosome dynamics and transcription *Genes Dev.* **24** 2031–42

Fakhouri W D, Ay A, Sayal R, Dresch J, Dayringer E and Arnosti D N 2010 Deciphering a transcriptional regulatory code: modeling short-range repression in the *Drosophila* embryo *Mol. Syst. Biol.* **6** 341

Garcia H G, Sanchez A, Kuhlman T, Kondev J and Phillips R 2010 Transcription by the numbers redux: experiments and calculations that surprise *Trends Cell Biol.* **20** 723–33

Gertz J, Siggia E D and Cohen B A 2009 Analysis of combinatorial *cis*-regulation in synthetic and genomic promoters *Nature* **457** 215–8

He X, Samee M A, Blatti C and Sinha S 2010 Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression *PLoS Comput. Biol.* **6** e1000935

Jaeger J *et al* 2004 Dynamic control of positional information in the early *Drosophila* embryo *Nature* **430** 368–71

Janssens H, Hou S, Jaeger J, Kim A R, Myasnikova E, Sharp D and Reinitz J 2006 Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster* even skipped gene *Nat. Genet* **38** 1159–65

Kaplan T, Li X Y, Sabo P J, Thomas S, Stamatoyannopoulos J A, Biggin M D and Eisen M B 2011 Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early *Drosophila* development *PLoS Genet* **7** e1001290

Li L M and Arnosti D N 2011 Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes *Curr. Biol.* **21** 406–12

Makeev V J, Lifanov A P, Nazina A G and Papatsenko D A 2003 Distance preferences in the arrangement of binding motifs and hierarchical levels in organization of transcription regulatory information *Nucleic Acids Res.* **31** 6016–26

McGhee J D and von Hippel P H 1974 Theoretical aspects of DNA–protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice *J. Mol. Biol.* **86** 469–89

Miller J A and Widom J 2003 Collaborative competition mechanism for gene activation *in vivo Mol. Cell Biol.* **23** 1623–32

Papatsenko D, Goltsev Y and Levine M 2009 Organization of developmental enhancers in the *Drosophila* embryo *Nucleic Acids Res.* **37** 5665–77

Rippe K 2001 Making contacts on a nucleic acid polymer *Trends Biochem. Sci.* **26** 733–40

Saiz L, Rubi J M and Vilar J M 2005 Inferring the *in vivo* looping properties of DNA *Proc. Natl Acad. Sci. USA* **102** 17642–5

Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U and Gaul U 2008 Predicting expression patterns from regulatory sequence in *Drosophila* segmentation *Nature* **451** 535–40

Teif V B 2007 General transfer matrix formalism to calculate DNA-protein-drug binding in gene regulation: application to $O_R$ operator of phage λ *Nucleic Acids Res.* **35** e80

Teif V B 2010 Predicting gene-regulation functions: lessons from temperate bacteriophages *Biophys. J.* **98** 1247–56

Teif V B, Ettig R and Rippe K 2010 A lattice model for transcription factor access to nucleosomal DNA *Biophys. J.* **99** 2597–607

Teif V B and Rippe K 2010 Statistical-mechanical lattice models for protein–DNA binding in chromatin *J. Phys.: Condens. Matter* **22** 414105

Yuan Y, Guo L, Shen L and Liu J S 2007 Predicting gene expression from sequence: a reexamination *PLoS Comput. Biol.* **3** e243

Yuh C H, Bolouri H and Davidson E H 1998 Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene *Science* **279** 1896–902

Zinzen R P, Senger K, Levine M and Papatsenko D 2006 Computational models for neurogenic gene expression in the *Drosophila* embryo *Curr. Biol.* **16** 1358–65