GENOMICS AND COMPUTATIONAL BIOLOGY Vol. 3 No. 2 (2017): e54

Short Communication

Mathematical Modelling of Promoter Occupancies in MYC-Dependent Gene Regulation

Uwe Benary[1,*], Elmar Wolf[2], Jana Wolf[1]

[1]Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Robert-Rössle-Straße 10, 13125 Berlin-Buch, Germany
[2]Department of Biochemistry and Molecular Biology, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany

*Correspondence: Tel: +49 30 9406 2465; Fax: +49 30 9406 2394; Email: uwe.benary@mdc-berlin.de

Received 2016-10-14; Accepted 2016-11-03

ABSTRACT

The human Myc proto-oncogene protein (MYC) is a transcription factor that plays a major role in the regulation of cell proliferation. Deregulation of MYC expression is often found in cancer. In the last years, several hypotheses have been proposed to explain cell type specific MYC target gene expression patterns despite genome wide DNA binding of MYC. In a recent publication, a mathematical modelling approach in combination with experimental data demonstrated that differences in MYC-DNA-binding affinity are sufficient to explain distinct promoter occupancies and allow stratification of distinct MYC-regulated biological processes at different MYC concentrations. Here, we extend the analysis of the published mathematical model of DNA-binding behaviour of MYC to demonstrate that the insights gained in the investigation of the human osteosarcoma cell line U2OS can be generalized to other human cell types.

KEYWORDS

MYC; mathematical modelling; cell type specificity; E-box

INTRODUCTION

The human *MYC* gene codes for a transcription factor called Myc proto-oncogene protein (in the following abbreviated with MYC) that is involved in a variety of cellular processes, including regulation of cell proliferation [1, 2]. Increased expression levels of MYC contribute to cancer development and are associated with tumour aggressiveness [3]. Several studies suggest that MYC binds to a specific DNA motif termed enhancer box (E-box), with a consensus sequence of CACGTG, and thereby regulates the transcription of specific target genes [4, 5]. More recent experiments using chromatin-immunoprecipitation (ChIP) and ChIP-sequencing experiments [6–9] demonstrated that MYC binds to almost all promoters, enhancers and intergenic sites with an open chromatin structure, independent of the presence of E-boxes [10, 11]. Despite this global DNA binding, MYC-dependent tumours seem to harbour a specific set of up- and down-regulated MYC target genes.

Two seemingly conflicting hypotheses have been proposed to explain these observations [12, 13]. One hypothesis, termed the general amplifier model [7, 8], suggests that MYC globally enhances transcription and that specific gene expression arises indirectly due to regulatory feedback and feedforward loops. The alternative hypothesis suggests that a large portion of MYC binding to DNA is non-productive with respect to transcriptional regulation resulting in a specific set of regulated genes [14, 15]. Global changes in mRNA levels in this model are the consequence of physiological and metabolic changes induced by specific sets of regulated genes.

In our recent study, we proposed that the divergent hypotheses can be reconciled in the new hypothesis that specific gene expression profiles arise from different affinities of MYC-DNA binding [11]. We used a mathematical modelling approach in combination with extensive experimental data to show that differences in affinity are sufficient to explain the distinct promoter occupancies observed in ChIP-sequencing experiments for physiological and tumour-specific MYC concentrations. Moreover, the promoter affinities estimated from experimental data determine the impact of MYC concentration changes on target gene expression and they allow a stratification of distinct MYC-regulated cellular processes at different MYC concentrations. We also suggested that interactions between MYC and promoter-bound factors may increase promoter affinities, which can serve as a molecular mechanism of context-specific modulation of MYC-dependent transcriptional responses of individual genes [13].

Our published model of MYC-DNA binding was parameterised to represent U2OS cells. It was comprehensively analysed with respect to the influence of its parameters, such as assumed affinities and proportion of heterochromatin, on the occupancy of E-boxes by MYC. Many other human cell types differ in several aspects from U2OS cells. For instance, their karyotype and/or the fraction of E-boxes in their genome may deviate due to single-nucleotide polymorphisms (SNPs), larger structural variations, segmental deletions or copy number variations, to name a few [16–19]. We here present an extended analysis of our published mathematical model of MYC-DNA-binding [11] to demonstrate that the insights gained in the investigation of U2OS cells may be generalized to other human cell types as well.

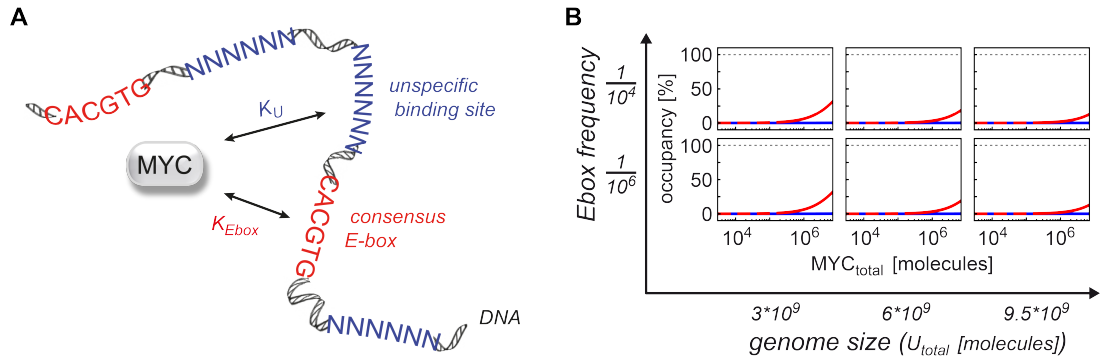

Figure 1: Extended analysis of the mathematical model of MYC-DNA-binding. A) Schematic representation of the mathematical model. A detailed mathematical description of the model is provided in [11]; in brief, MYC can reversibly bind to either consensus E-boxes or unspecific binding sites on the DNA. B) Impact of total MYC levels on occupancies of E-boxes (red curves) and unspecific binding (blue curves) sites shown for particular combinations of genome sizes and E-box frequencies.

METHODS

Mathematical modelling

We use the mathematical model 1 as described earlier ([11] and Appendix 1 therein). Briefly, MYC can reversibly bind to either consensus E-boxes or unspecific binding sites on the DNA. For convenience, we renamed the two published parameters K_{NNNNNN} and $[NNNNNN_{total}]$ as K_U and $[U_{total}]$ in our study. Steady-state solutions were calculated by setting all time derivatives in the model equations to zero and solving the resulting algebraic equation system numerically. All calculations were performed using Mathematica 10.2 (Wolfram Research, Inc.).

RESULTS

The published mathematical model describes the reversible association of MYC with either consensus E-boxes or unspecific DNA binding sites. It was parameterised to represent U2OS cells. Two kinds of parameters are considered in the model: (I) dissociation constants (K_{Ebox}, K_U) interpreted as inverse measures of affinity of MYC-E-box-binding and unspecific MYC binding to DNA, respectively (Figure 1A), and (II) total numbers of molecules and binding sites ($[MYC_{total}]$, $[Ebox_{total}]$, $[U_{total}]$).

In our analysis, we consider the dissociation constants (K_{Ebox}, K_U) as cell type independent biochemical properties and keep them fixed to their published values. The impact that changes of these dissociation constants have on occupancies was comprehensively explored in our earlier study [11]. In contrast to the dissociation constants, total numbers of molecules and binding sites are certainly cell type specific and are varied in our analysis.

MYC quantification in different human cancer cell lines yielded roughly 10^5 to 10^6 molecules per cell [7, 11]. A generally accepted estimate of mammalian transcription factors is 10^4-10^6 molecules per cell [20]. Hence, to cover all probable cellular conditions, we vary $[MYC_{total}]$ over a rather broad range in our analysis (Figure 1B).

We also consider different numbers of E-boxes and genome sizes than were assumed in case of U2OS cells ($9.3*10^5$ molecules and $9.5*10^9$ molecules, respectively). The total genome length of a stereotypic human cell is approximately 3 billion base pairs, which enumerates to about $3*10^9$ hexameric binding sites ($[U_{total}]$) in a haploid cell and about $6*10^9$ sites in a diploid cell (Figure 1B, panels in left column and middle column, respectively). In the human reference genome on average 1 consensus E-box is found in approximately every 10^4 hexameric binding sites (Figure 1B, panels in upper row), i.e. $[Ebox_{total}]=9.3*10^5$ molecules in U2OS cells.

An alternative, likewise plausible estimate of $[Ebox_{total}]$ can be derived from the notion that MYC regulates the expression of 15% of all human genes [21]. Given that the number of human protein-coding genes is in the order of $2*10^4$ [22] we can estimate about 3000 MYC regulated genes. For simplicity, let us now presume that each promoter of these genes harbours a consensus E-box, we can roughly estimate an average frequency of 1 consensus E-box in every 10^6 hexameric binding sites (Figure 1B, panels in bottom row).

Having thus confined the parameter space of our model to plausible ranges, we next explore the impact of total MYC levels on occupancies of E-boxes and unspecific DNA binding sites for different combinations of genome sizes ($[U_{total}]$) and E-box frequencies ($[Ebox_{total}]/[U_{total}]$). Our analysis demonstrates that for a given genome site changes in the frequency of E-boxes over orders of magnitude hardly affect the dependency of E-box occupancy on cellular MYC levels (Figure 1B; compare red curves in upper panels vs. bottom panels). It also shows for very high amounts of MYC ($>10^6$ molecules) that haploid cells can realise a visibly higher E-box occupancy compared to diploid cells or cells with even larger genomes such as U2OS cells. Note that for higher affinities of MYC to E-boxes, e.g. due to binding of promoter-bound factors, these changes of E-box occupancies can occur also at lower total MYC concentrations. In contrast, the occupancy of unspecific binding sites is hardly influenced by changes in total MYC level under all considered combinations of genome size and E-box frequency (Figure 1B; blue curves).

DISCUSSION

Most cells in the human body harbour a diploid karyotype. Cells may however differ in the number of promoters with transcriptional capacity, for instance by restricting the frequency of E-boxes available for MYC binding. Our analysis demonstrates that a change in the frequency of E-boxes in a cell has only minor impact on their occupancy by MYC. This may indicate that silencing of a large proportion of MYC target gene promoters, for instance by chromatin remodelling, hardly affects the occupancies of the remaining unsilenced promoters. Also preventing MYC binding to certain promoters by any other mechanism will not influence the remaining target gene promoters. The reason is that the impeded MYC molecules in such a scenario predominantly bind to the unspecific binding sites that are in vast excess over E-boxes. In fact, most MYC (more than 95% of total MYC) is bound to unspecific binding sites in all our model simulations. The observed low occupancy of unspecific binding sites is simply due to the large number of these sites compared to total MYC molecules. We conclude that cell type specific differences in total E-boxes hardly matter in the analysis of the model for a given genome size. Differences in genome size between haploid, diploid and U2OS cells seem to influence E-box occupancies at certain total MYC concentrations depending on the affinity of MYC. The presented results therefore demonstrate the robustness of the model behaviour with respect to parameterisation indicating that our previous insights and conclusions can be generalized to other human cell types as well.

ACKNOWLEDGEMENTS

This study was funded by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept to JW (01ZX1307F) and supported by a grant from the Deutsche Forschungsgemeinschaft to EW (Emmy Noether Research Group, WO 2108/1-1).

AUTHOR CONTRIBUTIONS

UB carried out model analysis and wrote the text. JW supervised the study. All authors read over and approved the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

MYC: Myc proto-oncogene protein (UniProtKB: P01106)
ChIP: chromatin-immunoprecipitation
SNPs: single-nucleotide polymorphisms
E-box: enhancer box

REFERENCES

1. Meyer N, Penn LZ. **Reflecting on 25 years with MYC**. Nature Reviews Cancer. 2008 Dec;8(12):976–990. doi:10.1038/nrc2231.
2. Conacci-Sorrell M, McFerrin L, Eisenman RN. **An Overview of MYC and Its Interactome**. Cold Spring Harbor Perspectives in Medicine. 2014 Jan;4(1):a014357. doi:10.1101/cshperspect.a014357.
3. Posternak V, Cole MD. **Strategically targeting MYC in cancer**. F1000Res. 2016 Mar;5:408. doi:10.12688/f1000research.7879.1.
4. Desbarats L, Gaubatz S, Eilers M. **Discrimination between different E-box-binding proteins at an endogenous target gene of c-myc.** Genes & Development. 1996 Feb;10(4):447–460. doi:10.1101/gad.10.4.447.
5. Prendergast G, Ziff E. **Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region**. Science. 1991 Jan;251(4990):186–189. doi:10.1126/science.1987636.
6. Fernandez PC. **Genomic targets of the human c-Myc protein**. Genes & Development. 2003 Apr;17(9):1115–1129. doi:10.1101/gad.1067003.
7. Lin C, Lovén J, Rahl P, Paranal R, Burge C, Bradner J, et al. **Transcriptional Amplification in Tumor Cells with Elevated c-Myc**. Cell. 2012 Sep;151(1):56–67. doi:10.1016/j.cell.2012.08.026.
8. Nie Z, Hu G, Wei G, Cui K, Yamane A, Resch W, et al. **c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells**. Cell. 2012 Sep;151(1):68–79. doi:10.1016/j.cell.2012.08.033.
9. Sabò A, Kress TR, Pelizzola M, de Pretis S, Gorski MM, Tesi A, et al. **Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis**. Nature. 2014 Jul;511(7510):488–492. doi:10.1038/nature13537.
10. Walz S, Lorenzin F, Morton J, Wiese KE, von Eyss B, Herold S, et al. **Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles**. Nature. 2014 Jul;511(7510):483–487. doi:10.1038/nature13473.
11. Lorenzin F, Benary U, Baluapuri A, Walz S, Jung LA, von Eyss B, et al. **Different promoter affinities account for specificity in MYC-dependent gene regulation**. eLife. 2016 jul;5:e15161. doi:10.7554/eLife.15161.
12. Levens D. **Cellular MYCro Economics: Balancing MYC Function with MYC Expression**. Cold Spring Harbor Perspectives in Medicine. 2013 Nov;3(11):a014233–a014233. doi:10.1101/cshperspect.a014233.
13. Wolf E, Lin CY, Eilers M, Levens DL. **Taming of the beast: shaping Myc-dependent amplification**. Trends in Cell Biology. 2015 Apr;25(4):241–248. doi:10.1016/j.tcb.2014.10.006.
14. Kress TR, Sabò A, Amati B. **MYC: connecting selective transcriptional control to global RNA production**. Nat Rev Cancer. 2015 Sep;15(10):593–607. doi:10.1038/nrc3984.
15. Sabo A, Amati B. **Genome Recognition by MYC**. Cold Spring Harbor Perspectives in Medicine. 2014 Feb;4(2):a014191. doi:10.1101/cshperspect.a014191.
16. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. **The Diploid Genome Sequence of an Individual Human**. PLoS Biol. 2007 Sep;5(10):e254. doi:10.1371/journal.pbio.0050254.
17. Junker J, van Oudenaarden A. **Every Cell Is Special: Genome-wide Studies Add a New Dimension to Single-Cell Biology**. Cell. 2014 Mar;157(1):8–11. doi:10.1016/j.cell.2014.02.010.
18. Voet T, Kumar P, Van Loo P, Cooke SL, Marshall J, Lin ML, et al. **Single-cell paired-end genome sequencing reveals structural variation per cell cycle**. Nucleic Acids Research. 2013 Apr;41(12):6119–6138. doi:10.1093/nar/gkt345.
19. Durbin RM, Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, et al. **A map of human genome variation from population-scale sequencing**. Nature. 2010 Oct;467(7319):1061–1073. doi:10.1038/nature09534.
20. Moran U, Phillips R, Milo R. **SnapShot: Key Numbers in Biology**. Cell. 2010;141(7):1262 – 1262.e1. doi:10.1016/j.cell.2010.06.019.
21. Gearhart J, Pashos EE, Prasad MK. **Pluripotency Redux — Advances in Stem-Cell Research**. New England Journal of Medicine. 2007 Oct;357(15):1469–1472. doi:10.1056/nejmp078126.
22. Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, et al. **Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes**. Human Molecular Genetics. 2014 Jun;23(22):5866–5878. doi:10.1093/hmg/ddu309.