Biochem. Soc. Symp. **73**, 67–75
(Printed in Great Britain)

7

Specialized transcription factories

Jon Bartlett, Jelena Blagojevic, David Carter, Christopher Eskiw, Maud Fromaget, Christy Job, Monee Shamsher, Inês Faro Trindade, Meng Xu and Peter R. Cook

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, U.K.

Abstract

We have previously suggested a model for the eukaryotic genome based on the structure of the bacterial nucleoid where active RNA polymerases cluster to loop the intervening DNA. This organization of polymerases into clusters – which we call transcription 'factories' – has important consequences. For example, in the nucleus of a HeLa cell the concentration of soluble RNA polymerase II is ~1 mM, but the local concentration in a factory is 1000-fold higher. Because a promoter can diffuse ~100 nm in 15 s, one lying near a factory is likely to initiate; moreover, when released at termination, it will still lie near a factory, and the movement and modifications (e.g. acetylation) accompanying elongation will leave it in an 'open' conformation. Another promoter out in a long loop is less likely to initiate, because the promoter concentration falls off with the cube of the distance from the factory. Moreover, a long tether will buffer it from transcription-induced movement, making it prone to deacetylation, deposition of HP1 (heterochromatin protein 1), and incorporation into heterochromatin. The context around a promoter will then be self-sustaining: productive collisions of an active promoter with the factory will attract factors increasing the frequency of initiation, and the longer an inactive promoter remains inactive, the more it becomes embedded in heterochromatin. We review here the evidence that different factories may specialize in the transcription of different groups of genes.

Introduction

A 'factory' is defined in The Oxford English Dictionary as "a building or range of buildings with plant for the manufacture of goods". We use the term

[1]To whom correspondence should be addressed (email peter.cook@path.ox.ac.uk).

to describe the sites of replication, transcription and repair of damaged DNA in eukaryotic cells, as each site contains many different polymerizing complexes working on many different templates [1–5]. We have also proposed a general model for the organization of all genomes in which the factories involved in transcription play a central role (Figure 1) [6,7]. When genes strung along a template are transcribed, active polymerases and transcription factor–DNA complexes cluster to form the factories, with a consequential looping of intervening DNA. Each factory would then be surrounded by a 'cloud' of loops, and strings of nucleosomes and factories (plus surrounding clouds) would constitute the major architectural motifs responsible for organizing the genome. Active polymerases do not track along their templates; they are bound to a factory, and act both as motors that reel in their templates and as one of the critical structural ties that maintain the loops. DNA within a loop is free to diffuse within the volume set by the length of the tether. As engaged polymerases and transcription factors constitute the molecular ties, loops inevitably appear and disappear as polymerases initiate and terminate, and the factors bind and dissociate.

Clustering ensures that the local concentration of polymerases and promoters in and around a factory is high, enabling them to interact efficiently. For example, HeLa cell nuclei contain a dispersed pool of RNA polymerase II present at a concentration of ~1 μM, but a fraction is found in transcription factories, where the local concentration is ~1000-fold higher [7]; as a result, few – if any – mRNAs are made outside these factories (e.g. [8]). A promoter lying close to a factory is likely to collide with a polymerase and initiate; moreover, when released at termination, it will still lie near a factory, and the movement and modifications (e.g. acetylation) accompanying elongation will leave it in an 'open' conformation. Another promoter out in a long loop is less likely to initiate, because the promoter concentration falls off with the cube of the distance from the factory. Moreover, a long tether will buffer it from transcription-induced movement, making it prone to deacetylation, and incorporation into heterochromatin. The context around a promoter will then be self-sustaining: productive collisions of an active promoter with the factory will attract the appropriate factors, which will inevitably increase the frequency of initiation; on the other hand, the longer an inactive promoter remains inactive, the more it becomes embedded in heterochromatin. Here we review the evidence that different factories might specialize in the transcription of different groups of genes.

Nucleolar transcription factories

Nucleoli provide an important precedent for the kind of specialization that we will consider [9]. Eukaryotic nuclei contain three different kinds of RNA polymerase (i.e. polymerases I, II and III); one of these – RNA polymerase I – is concentrated in nucleoli, where it transcribes a specific group of genes, the repeated ribosomal cistrons. Each nucleolus contains one or more clusters of polymerases at its core. Each is dedicated to both the synthesis of 45 S rRNA and the packaging of the resulting processed transcripts into ribosomes. The human loci encoding 45 S rRNA are carried on chromosomes 13, 14, 15, 21 and

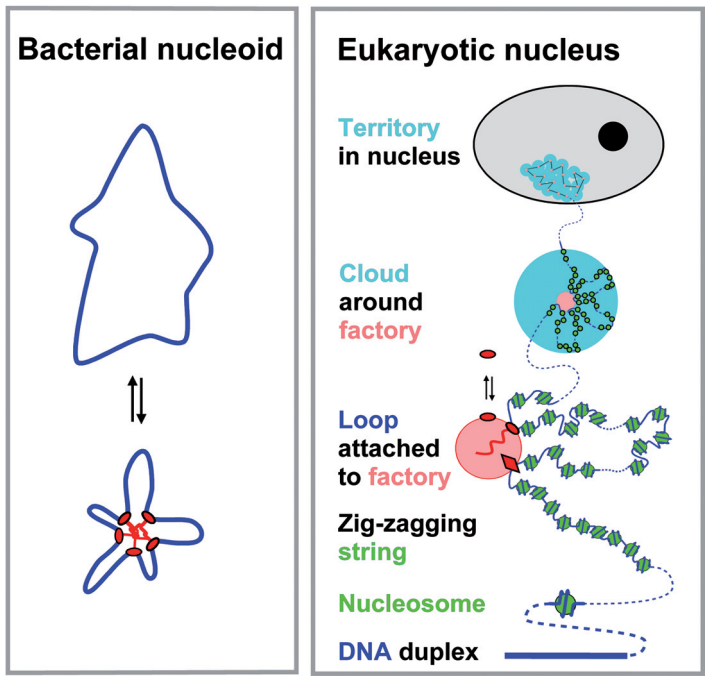

Figure 1 Models for genome structure. Left panel: In bacteria, transcription of the circular chromosome (top), followed by aggregation of polymerases (ovals) and transcripts (red lines), generates a looped structure (bottom) that is self-sustaining (as promoters in active genes now lie close to polymerases). Reproduced from [7]. Right panel: In eukaryotes (and specifically in a HeLa cell), DNA is coiled around the histone octamer, and runs of nucleosomes form a zig-zagging string. At the intermediate level in the hierarchy, this string is organized into loops (average contour length 86 kbp; range 5–200 kbp) by attachment to factories through transcription factors (diamond) and engaged RNA polymerases (ovals). Around 10–20 such loops (only a few are shown) form a 'cloud' around the factory, to give a structure equivalent to that of the bacterial nucleoid. Active polymerases do not track along their templates; they are bound to a factory and act both as motors that reel in their templates and as one of the critical structural ties that maintain the loops. Loops inevitably appear and disappear as polymerases initiate and terminate, and the factors bind and dissociate. Each factory contains one type of RNA polymerase (i.e. I, II or III) to the exclusion of others, and some factories are richer in certain transcription factors than others (and so are involved in the transcription of specific sets of genes). Individual components in the factory exchange continually with others in the soluble pool. A territory is formed by 50–200 successive clouds strung along the chromosome (the general path of DNA between clouds is shown). Modified from Cook [40], with permission of John Wiley & Sons, Inc.

22, with each locus carrying ~80 tandem repeats of 43 kbp containing the gene and an untranscribed spacer. These repeats appear as secondary constrictions in mitotic chromosomes – the NORs (nucleolar organizing regions). Inactive RNA polymerase I and its transcription factor UBF (upstream binding factor) are bound to some NORs, and – on exit from mitosis – these NORs fuse into one or more nucleoli. NORs lacking bound UBF and the polymerase remain inactive and are not initially incorporated into nucleoli, but they may be later if the (heterochromatic) centromeres to which they are linked associate with nucleolar heterochromatin.

Three distinct regions can be seen by electron microscopy: a granular component, in which are embedded one or more fibrillar centres and associated dense fibrillar component(s) [10,11]. The fibrillar centre contains high concentrations of polymerase I and UBF, and transcription takes place on its surface, as nascent Br-RNA is found in the enveloping dense fibrillar component. Nascent transcripts can be labelled by incubating permeabilized cells in Br-UTP. Newly completed transcripts are processed in the granular component to emerge as mature ribosomal subunits into the nucleoplasm. A typical nucleolar factory in a HeLa cell contains ~500 polymerases engaged on around four templates, with zones dedicated to storage, RNA synthesis and processing [10,12].

Nucleoplasmic factories containing either RNA polymerase II or III

Active RNA polymerase I is clearly contained in a specialized compartment, where it transcribes a particular group of genes, but what of the nucleoplasmic polymerases? As quantitative analysis of the numbers of nascent transcripts and sites had revealed that each site must contain many active polymerases, we can envisage two different types of organization. In the first, a cluster of polymerase II molecules might form one factory where all the enzymes work on class II transcription units; an analogous cluster of polymerase III molecules would work only on class III units. In the second, one factory might contain both types of enzyme to generate both class II and III transcripts. Three approaches suggest that the former applies, and that sites contain one type of active enzyme but not the other [8]. In the first, we compared site densities after inhibiting polymerase II with α-amanitin. If the two types of polymerase are intermingled within a site, and if most sites are detected (which was shown to be the case), we would expect inhibition of polymerase II to have little effect on site density. However, site density fell to one-fifth, consistent with one-fifth of the sites being dedicated to transcription by polymerase III. The second approach involved seeing if a polymerase lay next to transcripts made by the other enzyme. Although polymerases II and III often lay immediately next to their own transcripts, they rarely (if at all) lay next to those made by the other enzyme. However, such negative results are indecisive, especially when the approach is bedevilled by steric hindrance between immunolabelling probes. Convincing evidence against co-localization was obtained using a third approach that exploited this steric hindrance; cryosections of appropriately

labelled nuclei were preincubated with an antibody to see if it blocked access of a second. We found that an anti-(polymerase II) antibody blocked access to transcripts made by polymerase II, but not to polymerase III protein or its transcripts; conversely, an antibody against polymerase III blocked access to polymerase III transcripts, but not to polymerase II protein or its transcripts. Therefore the results of all three approaches are consistent with polymerases II and III being active in spatially distinct sites.

These results suggest that the two nucleoplasmic polymerases – like the nucleolar enzyme – are concentrated in their own dedicated factories. Quantitative analysis shows that each HeLa cell nucleus typically contains ~90 000 nascent transcripts, with ~15 000, ~65 000 and ~10 000 being made by polymerases I, II and III respectively [12]. Each of the ~8000 polymerase II factories contains around eight polymerases engaged on eight different templates, while each of the ~2000 polymerase III factories contains around five polymerases active on five different templates [8].

Complementary functions in one RNA polymerase II factory

The functions required to generate a 'standard' mRNA made by RNA polymerase II (e.g. capping, splicing and polyadenylation) were originally thought to occur independently, but it is now clear that they are interdependent [13]. Thus the C-terminal domain of the catalytic subunit of polymerase II – which in humans contains 52 heptad repeats with the consensus sequence YSPTSPS – forms a platform that integrates these functions. This enormous complex also proofreads newly made transcripts before going on to destroy faulty ones [14–16]. Proofreading involves the NMD (nonsense-mediated decay) pathway [17], and probably uses ribosomes to scan transcripts for inappropriately placed (i.e. premature) termination codons. Nascent peptides made during ribosomal scanning are found in nucleoplasmic factories, together with components of the translation and NMD machineries. As inhibiting transcription immediately inhibits this nuclear protein synthesis, the two processes must be tightly coupled. Moreover, the transcriptional, translational, NMD and degradative machinery (which destroys faulty transcripts and any misfolded peptides resulting from translational proofreading) all co-localize and co-immunoprecipitate; selected components (translational initiation factor eIF4E, ribosomal subunit S6, NMD factors UPF1 and UPF2) also co-purify with the catalytic subunit of the polymerase, probably by binding to its C-terminal domain. However, it remains unclear whether all the different machineries are bound at any given moment [18].

Further specialization of nucleoplasmic factories

Do nucleoplasmic factories specialize further, to transcribe particular gene sets? It seems that they might, as active sets associate with specific sites (much as active NORs associate with nucleoli). Some examples will now be discussed.

The OPT domain

PTF [PSE (proximal sequence element)-binding transcription factor] activates transcription of snRNA (small nuclear RNA) and related genes. Immunofluorescence shows that it is spread throughout the nucleoplasm of HeLa cells in small foci. In some cells, it is also concentrated in one – or a few – discrete foci (diameter ~1.3 μm) that appear during G_1 phase and disappear in S phase. Oct1 (a transcription factor that interacts with PTF), RNA polymerase II, TBP (TATA-binding protein) and Sp1 are also present in these foci. Each focus typically contains two or three transcription factories that can incorporate Br-UTP into nascent transcripts, so they were christened OPT (Oct1/PTF/transcription) domains. Significantly, a small region on chromosome 6 (i.e. band 6p21) that contains only ~30 Mbp DNA, plus chromosomes 6 and 7, associate with the domain significantly more than other chromosomes. This suggests that these domains may act like nucleoli to bring particular genes on specific chromosomes together to a region where the appropriate transcription and processing factors are concentrated, thereby facilitating the expression of those genes [19].

The β-globin transcription factory

Results from various methods – 3C (chromosome conformation capture), RNA TRAP (RNA tagging and recovery of associated proteins), ChIP (chromatin immunoprecipitation) and FISH (fluorescence *in situ* hybridization) – now provide good evidence for the existence of chromatin loops tethered through active transcription units to a 'hub' or factory (as in Figure 1), as well as providing suggestive evidence that one particular factory specializes in the transcription of a particular set of genes. The first three methods involve fixation, before analysis of which DNA sequences lie next to each other in three-dimensional space – after cutting, dilution and ligation (in 3C), or purifying complexes containing either a specified nascent transcript (in RNA TRAP) or protein bound to DNA (in ChIP).

The mouse *Hbb-b1* (β-globin) gene lies tens of kilobase-pairs away on chromosome 7 from its LCR (locus control region), and ~25 Mbp away from a gene (*Eraf*) encoding the α-globin stabilizing protein. 3C shows that *Hbb-b1* contacts the LCR and *Eraf* in erythroid nuclei (where all three are transcribed, e.g. see [20]), but not in brain nuclei, where all are inactive [21–24]. RNA TRAP confirms that *Hbb1* contacts the LCR [25]. DNA- and RNA-FISH coupled to immunolabelling also show that active *Hbb-b1* and *Eraf* are found together in sites rich in polymerase II; moreover, inactive alleles in the same cell are generally distant from such sites [23]. These factories – in which some of the most highly expressed genes in the erythroid lineage are concentrated – are larger than those in HeLa cells, and have been described as "active chromatin hubs" [21].

Note that we have argued elsewhere that a wide range of other LCRs, including those controlling the expression of α-globin, keratin 18, adenosine deaminase, growth hormone and MHC class II genes, all encode non-genic transcription units, and that transcription of these LCRs probably underlies their function [26]. Note also that where these new methods have been used to demonstrate looping in other cases, the tethering points also turn out to be

transcription units, in accordance with the model illustrated in Figure 1. For example, 3C shows that the two barriers flanking the *Drosophila* 87A7 heat-shock locus – *scs* and *scs′* – contact each other [27], and both barriers encode promoters [28]. Moreover, 3C and ChIP show that the mouse *H19* DMR (differentially methylated region) interacts with the *Igf2* DMR1 (but not DMR2) when *H19* is expressed from the maternal allele; conversely, it interacts with the *Igf2* DMR2 (but not DMR1) when *Igf2* is expressed from the paternal allele [29]. Both DMR1 and DMR2 are transcribed – the former encodes both sense and antisense transcription units, and the latter lies within *Igf2* [30]. (Unfortunately, we do not know whether the *H19* DMR is transcribed, but promoters do lie nearby [31].) So in all of these cases, it seems that transcription units contact each other – but only when they are active.

Stress granules

On heat shock, satellite III repeats on human chromosome 9q12 are transcribed and associate with a factory containing heat-shock factor 1 [32]. It becomes tempting to suggest that this factory contains the factors necessary for the transcription of this group of units.

Cajal bodies

Some class II genes encode transcripts with characteristic 3′ stem-loops but no introns or poly(A) tails (e.g. histone mRNAs, U1–U4, U11 and snRNAs). They are often found near Cajal bodies, unlike the class III genes that encode U6 snRNA and 5 S rRNA. This association depends on transcription by RNA polymerase II, as a promoter mutation or transcriptional inhibition prevent it [33–35]. Here, these genes encoding transcripts that are not processed in the usual way associate with ‘non-standard’ factories that lack the ‘normal’ processing machinery.

PML (promyelocytic leukaemia) bodies

Gene-dense regions of various human chromosomes, including the major histocompatibility locus, are transcribed at the surface of these structures [36,37]. Again, do these structures contain the appropriate high concentrations of transcription factors required for the transcription of this sub-set of genes?

Peri-nucleolar polymerase III factories

Yeast tRNA genes on different chromosomes associate with the nucleolar surface, and a point mutation in the promoter of one eliminates this association [38]. This suggests that polymerase III factories may associate with the surface of the nucleolus.

In all the described cases, particular groups of genes associate with a particular structure in order to be transcribed, presumably because that structure contains the appropriate machinery.

Conclusions

We have reviewed the evidence that the three different types of nuclear RNA polymerases (i.e. I, II and III) are each concentrated in their own dedicated factories, and that the different types of transcription unit (i.e. classes I, II and III) are only transcribed once they associate with a factory of the appropriate type. We have gone on to suggest that specialization is carried even further, with particular factories becoming enriched in particular factors so they become dedicated to the transcription of those units that require those factors. We imagine these structures to be dynamic, with the factors, polymerases and transcription units all exchanging continuously. Then, loop attachments mediated by transcription factors would persist for seconds, and those mediated by polymerases for the minutes required for the cycle of initiation, elongation and termination [39]. This turnover ensures that factories can evolve as conditions change; the chance transcription of a new unit would inevitably leave a new set of factors bound to the factory, and this would increase the chances that the same (or related) units might be transcribed there.

We thank the BBSRC, Cancer Research UK, the MRC and the Wellcome Trust for support.

References

1. Hozák, P., Hassan, A.B., Jackson, D.A. and Cook, P.R. (1993) Cell **73**, 361–373
2. Jackson, D.A., Hassan, A.B., Errington, R.J. and Cook, P.R. (1993) EMBO J. **12**, 1059–1065
3. Jackson, D.A., Balajee, A.S., Mullenders, L. and Cook, P.R. (1994) J. Cell Sci. **107**, 1745–1752
4. Jackson, D.A., Hassan, A.B., Errington, R.J. and Cook, P.R. (1994) J. Cell Sci. **107**, 1753–1760
5. Cook, P.R. (1999) Science **284**, 1790–1795
6. Cook, P.R. (1995) J. Cell Sci. **108**, 2927–2935
7. Cook, P.R. (2002) Nat. Genet. **32**, 347–352
8. Pombo, A., Jackson, D.A., Hollinshead, M., Wang, Z., Roeder, R.G. and Cook, P.R. (1999) EMBO J. **18**, 2241–2253
9. Shaw, P.J. and Jordan, E.G. (1995) Annu. Rev. Cell Dev. Biol. **11**, 93–121
10. Hozák, P., Cook, P.R., Schöfer, C., Mosgöller, W. and Wachtler, F. (1994) J. Cell Sci. **107**, 639–648
11. Shaw, P.J., Highett, M.I., Bevan, A.F. and Jordan, E.G. (1995) EMBO J. **14**, 2896–2906
12. Jackson, D.A., Iborra, F.J., Manders, E.M.M. and Cook, P.R. (1998) Mol. Biol. Cell **9**, 1523–1536
13. Maniatis, T. and Reed, R. (2002) Nature (London) **416**, 499–506
14. Iborra, F.J., Jackson, D.A. and Cook, P.R. (2001) Science **293**, 1139–1142
15. Iborra, F.J., Escargueil, A.E., Kwek, K.Y., Akoulitchev, A. and Cook, P.R. (2004) J. Cell Sci. **117**, 899–906
16. Iborra, F.J., Jackson, D.A. and Cook, P.R. (2004) J. Cell Sci. **117**, 5713–5720
17. Hilleren, P. and Parker, R. (1999) Annu. Rev. Genet. **33**, 229–260
18. Wetterberg, I., Zhao, J., Masich, S., Wieslander, L. and Skoglund, U. (2001) EMBO J. **20**, 2564–2574
19. Pombo, A., Cuello, P., Schul, W., Yoon, J.-B., Roeder, R.G., Cook, P.R. and Murphy, S. (1998) EMBO J. **17**, 1768–1778
20. Routledge, S.J. and Proudfoot, N.J. (2002) J. Mol. Biol. **323**, 601–611
21. Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F. and de Laat, W. (2002) Mol. Cell **10**, 1453–1465
22. Palstra, R.J., Tolhuis, B., Splinter. E., Nijmeijer, R., Grosveld, F. and de Laat, W. (2003) Nat. Genet. **35**, 190–194

23. Osborne, C.S., Chakalova, C., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W. and Fraser, P. (2004) Nat. Genet. **36**, 1065–1071
24. Patrinos, G.P., de Krom, M., de Boer, E., Langeveld, A., Imam, A.M., Strouboulis, J., de Laat, W. and Grosveld, F.G. (2004) Genes Dev. **18**, 1495–1509
25. Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.-f. and Fraser, P. (2002) Nat. Genet. **32**, 623–626
26. Cook, P.R. (2003) J. Cell Sci. **116**, 4483–4491
27. Blanton, J., Gaszner, M. and Schedl, P. (2003) Genes Dev. **17**, 664–675
28. Avramova, Z. and Tikhonov, A. (1999) Trends Genet. **15**, 138–139
29. Murrell, A., Heeson, S. and Reik, W. (2004) Nat. Genet. **36**, 889–893
30. Moore, T., Constancia, M., Zubair, M., Bailleul, B., Feil, R., Sasaki, H. and Reik, W. (1997) Proc. Natl. Acad. Sci. U.S.A. **94**, 12509–12514
31. Drewell, R.A., Arney, K.L., Arima, T., Barton, S.C., Brenton, J.D. and Surani, M.A. (2002) Development **129**, 1205–1213
32. Jolly, C., Metz, A., Govin, J., Vigneron, M., Turner, B.M., Khochbin, S. and Vourc'h, C. (2004) J. Cell Biol. **164**, 25–33
33. Callan, H.G., Gall, J.G. and Murphy, C. (1991) Chromosoma **101**, 245–251
34. Frey, M.R., Bailey, A.D., Weiner, A.M. and Matera, A.G. (1999) Curr. Biol. **9**, 126–135
35. Jacobs, E.Y., Frey, M.R., Wu, W., Ingledue, T.C., Gebuhr, T.C., Gao, L., Marzluff, W.F. and Matera, A.G. (1999) Mol. Biol. Cell **10**, 1653–1663
36. Shiels, C., Islam, S.A., Vatcheva, R., Sasieni, P., Sternberg, M.J.E., Freemont, P.S. and Scheer, D. (2001) J. Cell Sci. **114**, 3705–3716
37. Wang, J., Shiels, C., Sasieni, P., Wu, P.J., Islam, S.A., Freemont, P.S. and Sheer, D. (2004). J. Cell Biol. **164**, 515–526
38. Thompson, M., Haeusler, R.A., Good, P.D. and Engelke, D.R. (2003) Science **302**, 1399–1401
39. Kimura, H., Sugaya, K. and Cook, P.R. (2002) J. Cell Biol. **159**, 777–782
40. Cook, P.R. (2001) Principles of Nuclear Structure and Function, J. Wiley and Sons, New York