

Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression

David A. Hume

The phenotype of individual hematopoietic cells, like all other differentiated mammalian cells, is determined by selective transcription of a subset of the genes encoded within the genome. This overview summarizes the recent evidence that transcriptional regulation at the level of individual cells is best described in terms

of the regulation of the probability of transcription rather than the rate. In this model, heterogeneous gene expression among populations of cells arises by chance, and the degree of heterogeneity is a function of the stability of the mRNA and protein products of individual genes. The probabilistic nature of transcriptional

regulation provides one explanation for stochastic phenomena, such as stem cell lineage commitment, and monoallelic expression of inducible genes, such as lymphokines and cytokines. (Blood. 2000;96:2323-2328)

© 2000 by The American Society of Hematology

Introduction

Even the simplest organisms are able to modify the expression of certain genes in response to environmental signals such as nutrient levels. In complex multicellular organisms, selective gene expression is absolutely required for cellular differentiation and organogenesis and for homeostasis. Much of this regulation occurs at the level of transcription initiation; transcriptional regulatory proteins bind to the promoters of appropriate target genes and increase or decrease the amount of mRNA that is produced. Transcriptional regulation could be viewed as an analog process, akin to depressing the accelerator on a car, or a digital process, like switching on a light. In either a digital or an analog model, an increase in transcription in a single cell when a stimulus is added might occur as a direct and predictable response to a stimulus or in a probabilistic manner (the stimulus increasing the probability of response within a given time). The language of gene regulation generally assumes an analog model and direct causation; a transcription-activating factor is said to increase the rate of transcription. In this essay, I will present some of the evidence that transcription is actually a digital process, and I will argue that it is more meaningful to talk about the probability and frequency of transcription rather than the rate. In the second half of the essay, I will examine how such a view of transcription can change the way we interpret studies of inducible gene expression, cellular heterogeneity, and lineage determination using hematopoiesis and activation of T cells as examples.

bacterial lipopolysaccharide (LPS), increases the level of mRNA encoding a protease inhibitor, plasminogen activator inhibitor type 2 (PAI-2). The amount of detected PAI-2 mRNA in the cell population was a function of LPS concentration over quite a broad range. This increase was shown to result from increased transcription initiation using a nuclear run on transcription assay.¹ We concluded that LPS induces PAI-2 mRNA transcription, the implication being that each cell starts transcribing the gene more rapidly. However, using single-cell assays, we showed that the effect of increasing LPS concentration was actually to increase the number of cells with high levels of PAI-2 rather than the level of expression in each and every cell in the population. One might argue that each cell has its intrinsic activation threshold for LPS, but other genes, such as tumor necrosis factor (TNF)- α (which is regulated postranscriptionally in the macrophage cell line used), were induced in all the cells. In subpopulations of the macrophage cell line used in these studies, the frequency of expressing cells varied. Hence, PAI-2 mRNA production occurs in a stochastic manner, with a frequency (probability) that is determined by the strength of the LPS signal and the state of the cell. We call this a digital model, because it implies that the transcription apparatus exists in "on" and "off" states and that regulation involves switching between those states.

Biochemical studies of transcription initiation support a digital mechanism. The assembly of the transcription apparatus on a DNA template is unstable on a short time scale. Kadonaga² showed that the complex transcription apparatus must be reassembled between each "round" of transcription. Reassembly of a preinitiation complex occurred in vitro with a half-time of 3 minutes; thereafter, initiation occurred in seconds. It is important to recognize that the transcription machinery in such in vitro transcription systems is in large excess. Despite this excess, Bral et al³ found that only a small proportion of the templates in the assay were successful in forming

The digital process of transcription initiation

Transcription happens!

In a recent study of transcriptional regulation in macrophages, we used Northern blot analysis to show that the archetypal activator,

From the Departments of Biochemistry and Microbiology/Parasitology and the Centre for Molecular and Cellular Biology, University of Queensland, Queensland, Australia.

Submitted August 9, 1999; accepted April 18, 2000.

Supported by the National Health and Medical Research Council and the Queensland Cancer Fund. The Centre for Molecular and Cellular Biology is a Special Research Centre of the Australian Research Council.

Reprints: David A. Hume, Department of Biochemistry, University of Queensland, Q4072 Queensland, Australia; e-mail: d.hume@cmcb.uq.edu.au.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2000 by The American Society of Hematology

an active preinitiation complex. They propose that templates partition in a binary manner into active and inactive complexes. In a single cell, with only 2 DNA templates available for each specific gene, the formation of 2 inactive complexes would mean that the transcription of that gene is completely switched off. The question is, how long does it take, within an intact cell, to detach and disassemble a failed complex and try again?

The answer may be, quite a long time. The availability of antibodies against active phosphorylated RNA pol II, as well as other transcription factors or approaches to identifying nascent transcripts, has made it possible to localize the sites of transcription in the nucleus.^{4,5} Transcription seems to occur in specific physical structures within the nucleus, referred to as transcription factories.^{4,6} The number of sites of active pol II-mediated transcription in the nucleus was estimated to be approximately 2500.^{4,7} If this is the case, only a subset of protein-encoding genes is likely to be actively transcribed at any time. Aside from restrictions on the absolute availability in numerical terms, there is also evidence for functional compartmentation of transcription, processing and trafficking in the mammalian nucleus (reviewed in ⁸). Both RNA pol II and the splicing and processing factors required for co-transcriptional mRNA processing appear to be concentrated in discrete nuclear domains, sometimes referred to as speckles, of which there may be as few as 20 to 40 per nucleus. Translocation away from those domains is correlated with the activation of transcription.⁹⁻¹¹ Exactly how a DNA template identifies itself as a target for association with these components is beyond current understanding, but one can imagine the additional constraints on the probability of this event if RNA pol II templates are physically separated from the transcription apparatus.

A digital mechanism of transcriptional activation has been confirmed using single-cell imaging technology to study the process of transcription of β -actin, a gene commonly regarded as a control in studies of transcriptional regulation and one of the more abundant mRNAs and proteins in most cells.¹² Few of the serum-starved fibroblasts were found to be actively transcribing β -actin. The appearance of nascent transcripts on the 2 β -actin alleles was apparently activated in most cells within 5 minutes of serum addition to starved cells, but asynchronously. Once individual actin templates became activated in the serum-starved cells, the number of mRNA molecules per DNA template increased to a peak of 30 per template at 15 minutes, then decayed rapidly because of an abrupt cessation of initiation. At the peak of initiation, approximately 4 transcripts per minute were initiated on individual templates. This is more frequent than one might anticipate from the studies in vitro. One explanation is that once a successful preinitiation complex has been formed, reinitiation occurs with much higher probability. There is evidence that reinitiation can be independently regulated by transcriptional activators.¹³⁻¹⁶

As an alternative to direct visualization of events on individual templates within the nucleus, several groups have performed single-cell analyses of reporter gene expression. This approach introduces an additional level of complexity, because the relationship between the number of cells with detectable reporter gene product is clearly a function of the stability of the mRNA and protein and of transcriptional activity (see below). We first performed single-cell analyses using the HIV-1-LTR driving lacZ, stably transfected into RAW264 macrophages.¹⁷ In the absence of the Tat *trans*-activator, the frequency of strongly lacZ-positive cells was as low as 10^{-4} in cloned stable transfectants. Even assuming a short half-life for lacZ protein and mRNA, this implies that

transcription occurs infrequently indeed. In the presence of Tat, the frequency increased to approximately 10^{-2} and was increased still further by macrophage-activating stimuli. Fiering et al¹⁸ performed a similar study in the human Jurkat T-cell line and found that clonal lines with stably integrated lacZ reporters, driven by the IL-2 promoter or by multimerized *cis*-acting elements (κ B, NFAT-1), gave a bimodal distribution of lacZ activity. Accumulation of transcription factors in the nuclei of activated cells was correlated with transition between the lacZ-expressing and nonexpressing "states." Ko et al¹⁹ examined the expression of lacZ driven by steroid hormone-responsive elements and also found a bimodal distribution of reporter gene expression. The most striking example of such stochastic gene expression comes from recent studies using a nuclear lacZ reporter gene in muscle. In multinucleated myotubes the product of this transgene was localized specifically to the nucleus of origin, individual nuclei within a myotube either expressed the gene or did not, and the proportion of positive nuclei varied with differentiation state.²⁰

Evidence that gene expression is best described in terms of probability has also been provided from real-time imaging experiments. White et al²¹ described a system for determining the activity of luciferase reporter genes in transfected HeLa cells. They observed directly that the HIV-1-LTR and cytomegalovirus promoters shuttle on and off in individual cells. Subsequently, Takasuka et al²² used the same approach to study the regulation of the prolactin promoter in individual cells from a cloned pituitary cell line. Again, a profound variation without obvious pattern was observed between individual cells responding to stimuli that activate promoter activity.

Hence, a large body of evidence indicates that transcription is a digital process by which individual DNA templates exist in an off state and the likelihood of switching to the on state is regulated. If we accept that premise, we must use a new language to describe transcription. The production of mRNA occurs in *pulses*. The *mean frequency* of pulses is the major determinant of mRNA production and is determined by the *probability* of formation of a preinitiation complex. The average number of mRNA molecules produced in each pulse (which could be referred to *mean amplitude* of a pulse) is determined by the stability of the preinitiation complex, the *probability* of formation of a dead-end preinitiation complex in each successive round of reinitiation, or both. The *mean amplitude*, judging from the studies of β -actin, may be in the order of hundreds of transcripts. Because of the probabilistic basis of the model, this figure should display a normal distribution even among individual events occurring on one template and could be a target for regulation. Assuming that the powerful β -actin promoter is at the upper end of the spectrum of pulse sizes, an mRNA pulse might be measured in single digits for "weaker" promoters. This model might also be referred to as a quantal model. The distinction between a digital model of transcription and an analog model is conceptually similar to the difference between Newtonian and quantum mechanics, and it offers similar intellectual challenges.

Regulation of transcription: probabilities multiply!

Broadly speaking, the probability of transcription initiation can be regulated at 2 levels. Sequestration of genes into inactive chromatin, methylation, or other modifications can create a situation in which the gene is not available at all to the basal transcription apparatus (reviewed in ²³). Such sequestration is itself a stochastic process that is regulated by *cis*-acting elements in the vicinity of the gene. The effect of regulatory elements on the probability of transcriptional silencing is best demonstrated in the phenomenon of

transgene variegation.²⁴⁻²⁶ In genes that are “open for business,” the preinitiation complex must be assembled, involving TATA-binding protein and numerous accessory proteins that make up the basal transcription machinery. For this to occur the nucleosome structure in the vicinity of the active promoter must be disrupted to allow access, a process that is highly regulated.^{23,27-30} The discussion above emphasizes the probability of formation of a preinitiation complex as the major determinant of the frequency of transcriptional pulses on DNA templates. Although there is evidence for the regulation of reinitiation by some activators, many studies report that the addition of classical transcriptional activator proteins to an *in vitro* transcription assay causes transcriptional activation only if they are added before assembly of the preinitiation complex; that is, they modify the probability of the successful formation of such complexes.^{3,14,15,31-41}

The *cis*-acting elements recognized by transcription factors are commonly grouped in the vicinity of proximal promoters or enhancers. As predicted from the *in vitro* actions of transcription factors, compound elements such as enhancers act to increase the probability of transcription in intact cells.^{42,43} Within such enhancers, individual DNA-binding proteins may bind to each other and to DNA or may display interdependent binding activities. One relevant example is the complex IL-2 enhancer, where the binding or mutation of any *cis*-acting element abolishes binding of transcription factors to all the other sites.⁴⁴ The binding of transcription factors to intact chromatin, as opposed to naked DNA, may, in fact, be inherently cooperative.⁴⁵ There is also growing evidence for crucial roles for co-activator (and co-repressor) proteins in linking the binding of transcription activators to their individual response elements to the formation of an active preinitiation complex³⁹ or reinitiation on the same template.¹³ Each of these kinds of interaction can lead to genuine cooperativity or synergism in which transcription does not occur at all unless the entire machinery is available. However, the simple observation that 2 stimuli act multiplicatively is commonly taken as evidence of synergism, with an inferred mechanistic basis (eg, some form of protein-protein interaction).⁴⁶ An important corollary of the probabilistic view of transcription is that combinations of independent elements (or signals) generate sigmoidal dose response curves inherently because probabilities multiply. A model of a hypothetical “probability-driven” promoter is presented in Figure 1. Based on our experience with the HIV-1-LTR, a minimal promoter might have an intrinsic probability of transcription of one event in 10^4 templates per hour. In the model, each enhancer element occupied by a transcription factor produces a 5-fold increase in transcription probability (again, a realistic number based on our own experiences of transcriptional activators). Hence, if any 6 of those elements is occupied (say in response to an extracellular signal), the probability increases 5^6 -fold, so that most cells in the population will produce at least one pulse of transcripts within a 1-hour period. According to the probabilistic model, simple multiplicative actions of combinations of signals at any level in a transcriptional regulatory pathway can actually be taken as evidence against any direct interaction between them.

Biologic implications of transcription probability

Importance of mRNA stability

Each time a cell makes a pulse of a specific mRNA, it accumulates in the cytoplasm and is translated. Eventually the mRNA and

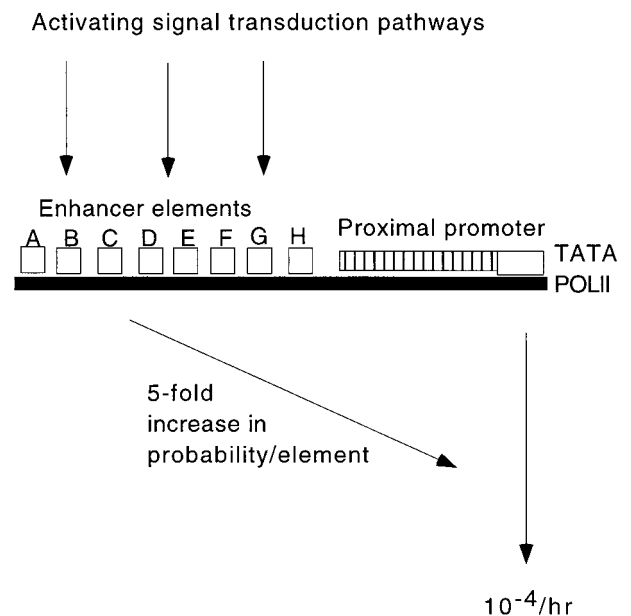


Figure 1. Probabilistic activation of a typical promoter. The diagram shows a typical inducible eukaryotic promoter. The basal promoter of the gene contains recognition sequences for abundant nuclear proteins such as Sp1, so that if the gene is in active chromatin, the elements are occupied. The minimal promoter ensures that there is a finite intrinsic low probability of recruitment of TATA-binding protein, and correct assembly an initiation complex including RNA polymerase II (POLII). The cell receives multiple independent signals simultaneously. Some of these may be autocrine, including products of the gene itself. Each signaling pathway increases the frequency with which the distal elements (A-H) are occupied. Each occupied site multiplies the probability of a pulse of transcript so that their effects in combination increase the frequency of transcription exponentially. Activation to a probability of one event per hour would be sufficient to ensure that in each cell at least one allele produces a pulse of transcript within a typical time frame of observation. Assuming absolutely independent regulation of each pathway, failure to activate any one of the *cis*-acting elements (A-H) would cause at least a 5-fold reduction in mRNA production. At a single-cell level, a probability of 0.2/gene per hour means that only 16% of cells would produce a pulse of the mRNA in a 1-hour time frame (because 4% will produce a second pulse).

protein decay. So, the time between pulses of transcription is characterized by a decay profile. If the mRNA and protein products are stable, the range of expression of individual genes in a population of cells will be small even if the gene is transcribed infrequently. For example, despite the absence of detectable active transcription, the level of β -actin mRNA in serum-starved fibroblasts was estimated at 500 ± 200 molecules per cell, rising to approximately 1500 after a pulse with serum.¹² If the half-life of β -actin mRNA is of the order of 10 hours, only occasional pulses of hundreds of mRNA molecules would be required to maintain that differential in the serum-stimulated steady state. The importance of mRNA stability is illustrated in another study by Kringstein et al,⁴⁷ who sought to examine the relation between transcription factor concentration and reporter gene production. The model system involved the use of a tetracycline-inducible system and a green fluorescent protein (GFP) reporter. Addition of activator was found to cause a step-wise dose-dependent increase in expression per cell, when assayed by flow cytometry. Their finding does not argue against an all-or-nothing process of transcriptional activation on single templates. It means simply that the GFP reporter is stable and accumulates in the cell in proportion to the number of rounds of transcription that occur within the time frame studied.

Conversely, a much broader distribution of levels of expression, approaching a true bimodality, will be observed if the mRNA and protein products of a gene have half-lives significantly shorter than

the average time between rounds of transcription. In these circumstances, the level of mRNA and protein in single cells becomes a function of the time that has elapsed since the previous pulse of mRNA. Clearly, the implication is that differential gene expression between “subpopulations” need not reflect some specific specialization; if the gene product has a short half-life, it just happens!

Stochastic regulation in hematopoiesis

A major issue in stem/progenitor cell fate determination in hematopoiesis concerns whether regulatory factors “instruct” target cells to commit to particular blood cell lineages or “select” cells that have already chosen a particular path that includes expression of the lineage-restricted receptor. There is a considerable body of evidence implying that lineage decisions in hematopoiesis arise probabilistically and that growth factors act to promote the survival and growth of cells in which the “decision” has already been made.^{48,49} A direct demonstration is found in a transgenic mouse expressing the human granulocyte-macrophage colony-stimulating factor (GM-CSF)-R, wherein human GM-CSF directed differentiation of a wide range of hematopoietic lineages.⁵⁰ This view point does not argue that colony-stimulating factors, or other regulators in the hematopoietic environment, have no function in lineage determination. The key distinction is between direct and predictable causation and the regulation of probability; a specific colony-stimulating factor may (and probably does) influence the likelihood that an individual target cell will follow a particular differentiation pathway.

The digital model of transcriptional regulation implies that individual genes can shuttle on and off in real time. If the gene products of key genes in hematopoietic commitment are unstable and transcribed sufficiently infrequently, then random pulses of mRNA could provide an explanation for heterogeneous expression and stochastic behavior. Stem cell “commitment” might arise through the co-expression, by chance, of combinations of lineage-specific transcription factors and hematopoietic growth factor receptor(s). At any early stage of commitment, any growth factor may suffice because it is the transcription factors that determine the cellular phenotype. Subsequently, the transcription factors will increase the probability of expression of lineage-restricted growth factor receptors (the promoters of which commonly contain binding sites for lineage-restricted transcription factors), further reinforcing the commitment event. Such a mechanism could only operate if most of the genes required for hematopoietic lineages were in open chromatin in stem cells and available for transcription, albeit with relatively low probability. The prediction of such a model is that stem cells and early “committed” progenitor cells would be extremely heterogeneous in gene expression. Iscove’s group⁵¹ developed single-cell-polymerase chain reaction approaches to identify mRNA expression in single cell-cloned hematopoietic progenitor cells. Their data revealed a fundamental randomness to co-expression of lineage-specific genes in specific lineage-committed colony-forming cells. Similarly, Hu et al⁵² found evidence of a so-called promiscuous phase of multi-lineage gene expression in which myeloid and erythroid lineage genes, including growth factor receptors, are co-expressed.

The most common assay of the phenomenon of hematopoietic commitment, the soft-agar cloning assay, is obviously binary in nature; hematopoietic growth factors are assayed based on the number of colonies they induce, and colonies are arbitrarily defined in size. A purely probabilistic model based on the chance expression of receptors predicts that the addition of more than one factor will have an additive effect on colony number. Remarkable cocktails of factors (IL-1, IL-3, IL-6, SCF, GM-CSF, CSF-1, and so on) are indeed added to colony assays to maximize the number of colonies formed.⁴⁹ Recently,

McKinstry et al⁵³ examined directly the expression on sorted hematopoietic stem/progenitor cell pools of receptors for a wide range of such factors. As predicted, they observe considerable heterogeneity in the percentage of cells labeled and the number of receptors per cell. They propose to determine whether such subpopulations have distinctive functional properties. If the subpopulations are “snapshots” of co-expression of certain genes arising by chance, such an approach may not be productive.

Inducible gene expression in lymphocytes

Several of the reporter genes discussed in the first section represent models of highly inducible genes. Differentiated cell types must maintain such genes in open chromatin but only express the gene product in response to an appropriate external stimulus. In response to that stimulus, they must rapidly increase the probability of transcription of the gene. One well-studied example of inducible gene regulation occurs in stimulated T lymphocytes. Activated T cells produce numerous different lymphokines. The prevailing paradigm is that subsets of activated T cells, termed Th1 and Th2 cells, produce different sets of lymphokines that polarize the immune response toward cell-mediated or humoral effector mechanisms, respectively. There has been a concerted effort to identify markers that define the Th1 and Th2 subsets of cells.⁵⁴ Of course, the tendency for 2 genes to be co-expressed, at both a single cell and a population level, can be greater than the product of 2 transcription probabilities. IL-4 and other Th2 lymphokines will be more likely to be produced together than predicted by chance if the genes that encode them share *cis*-acting elements. The drive toward co-expression can be amplified if the products of one gene, such as IL-4, activate signaling pathways that increase the probability of expression of other Th2-associated genes. Nevertheless, single-cell analysis has provided no evidence to support an absolute Th1/Th2 dichotomy.⁵⁵⁻⁵⁸ In fact, in studies of T cells activated in a T-cell receptor transgenic mouse, co-expression of any 2 lymphokine genes by individual T cells appeared to be the exception.⁵⁹⁻⁶¹ Bucy et al⁶² noted that during the development of so-called Th0 cell clones from such mice, each cytokine mRNA exhibited its own characteristic expression profile, and the major effect of increasing antigen dose was to “recruit” additional mRNA-expressing cells. Even in model systems in which stimuli were chosen to polarize the T-cell response selectively in the Th1 or Th2 direction, and in which selective cytokine production was demonstrable at the population level, single-cell mRNA analysis revealed that each gene is expressed with its own independent probability.⁵⁶ If this heterogeneity simply reflects the probability of assembly of inducible genes in active chromatin or assembly of the transcription complex once this has occurred, there is an obvious corollary. Unless the probability is very high, only one allele of any gene is likely to be transcribed in any one cell. Indeed, that prediction has recently been confirmed with the finding that inducible IL-2 and IL-4 production can be monoallelic in individual T cells.^{63,64} Similarly, in transgenic mice in which 1 allele of the IL-2 gene was replaced with a GFP reporter gene by homologous recombination, only a subset of individual activated T cells expresses the GFP reporter.⁶⁵ Bix and Locksley⁶⁶ confirmed the prediction that the pattern is random by identifying both bi-allelic and monoallelic expression of IL-4 among CD4-positive T-cell clones.

I predict that the availability of probes that distinguish the mRNA products of each allele in heterozygous individuals will show monoallelic expression to be a general phenomenon in inducible genes that has nothing to do with specific regulation. An interesting consequence of monoallelic expression arises when one

of the alleles is dysfunctional. The phenomenon of haploinsufficiency in genetic disease is generally considered to reflect a situation in which 50% of the normal level of gene product is insufficient to carry out normal function. Others have recognized^{67,68} that at a single-cell level, haploinsufficiency can mean complete insufficiency if the active allele is sequestered into inactive chromatin, or if the gene product is relatively unstable and the gaps between rounds of transcription are extended.

How does biology cope with transcriptional uncertainty?

Because transcriptional regulation is a process that underlies every aspect of eukaryotic biology and development, an acceptance of probabilistic basis for the process begs questions as to how order of any kind can be achieved. As noted by McAdams and Arkin,⁶⁹ stochastic patterns of gene expression can produce predictable outcomes with respect to particular alternative pathways. The uncertainty of transcription in eukaryotes is partly overcome by having 2 alleles at each locus and by the redundancy that seems particularly prevalent among transcription regulatory proteins in higher organisms. It is also overcome by having self-amplifying autocrine loops that are remarkably prevalent in systems such as the activation of macrophage cytokine production.⁷⁰ The certainty of activation of the gene in Figure 1 becomes much greater if each of the transcription factors also increases the probability of binding of the others (by protein-protein mechanisms or by increasing the probability of expression of other transcription factor genes) and if the gene product acts through autocrine/paracrine mechanisms to increase the frequency of activation of other alleles of the same gene. Finally, in real promoters, the number of transcription control elements can be remarkable. For example, in the inducible urokinase plasminogen activator gene in macrophages, we have described highly conserved transcription control elements extending up to 8 kb 5' of the transcription start site.^{71,72} If it simply had to be transcribed, the model gene in Figure 1 could have 20 control elements, so that occupation of any 10 of them would make transcription a certainty in biologic time.

Uncertainty could have its advantages. For example, if the mRNA and protein encoding a specific cell surface receptor is unstable, the level of receptor on each cell in the population will be heterogeneous. If a particular biologic response occurs in response to threshold level of occupied receptor, the number of responding cells will increase with agonist concentration. At another level, if the receptor interaction gives a quantitative signal, the fact that

each gene will have its own unique transcription probability dictated by the *cis*-acting elements it contains will dictate that each gene have its own unique dose-response curve to agonist. There are numerous examples in developmental biology that show concentration-dependent cell activation is required to allow a cell to determine its position in a gradient of a diffusible regulator. One example from cytokine regulation is the induction of TNF- α by LPS. A recent study used flow cytometry to demonstrate that the increase in cytokine production in blood monocytes treated with increasing doses of LPS is mainly caused by an increase in the number of cells responding.⁷³ We have reported that acute down-modulation of the CSF-1 receptor from the macrophage surface by macrophage activators is also all-or-nothing and that dose-response curves reflect the number of cells responding.⁷⁴ In a separate study we found that individual target genes are each induced by different doses of LPS and with different time courses in both primary macrophages and a macrophage cell line.¹ Random gene expression in cells of the innate and acquired immune systems could yield an extraordinary repertoire of potential effector cells to deal with every possible pathogen.

Conclusion

I have argued that transcription initiation in higher eukaryotes occurs with a relatively low frequency in biologic time and that the process is regulated in a probabilistic manner. In this model, the stability of the mRNA and protein products of a gene becomes crucially important. Some predictions of this model have been confirmed in hematopoietic cells; wherever single-cell (or single-allele) analysis is performed, the expression of each gene is seen to be regulated stochastically. The model changes the way we interpret the heterogeneity of cells in hematopoietic/immune systems. It may provide an explanation for the remarkable plasticity of stem cells,⁷⁵ which is one of the most exciting areas of current biology.

Acknowledgments

I thank Professor Alan Wolffe (National Institutes of Health, Bethesda, MD), Professor Anne Kelso (Queensland Institute for Medical Research, Brisbane, Australia), and Dr Jerry Molitor (University of Washington, Seattle, WA) for critical reading and helpful comments. I also thank the reviewers of *Blood*, who decimated the first version of this manuscript, hopefully to good effect.

References

- Costelloe EO, Stacey KJ, Antalis TM, Hume DA. Regulation of the plasminogen activator inhibitor type 2 gene in murine macrophages: demonstration of a novel pattern of responsiveness to bacterial endotoxin. *J Leukoc Biol*. 1999;66:172-183.
- Kadonaga JT. Assembly and disassembly of the *Drosophila* RNA polymerase II complex during transcription. *J Biol Chem*. 1990;265:2624-2631.
- Bral CM, Steinke JW, Kang CJ, Peterson DO. RNA polymerase II transcription complex assembly in nuclear extracts. *Gene Expr*. 1998;7:191-204.
- Iborra FJ, Pombo A, Jackson DA, Cook PR. Active RNA polymerases are localized with discrete transcription "factories" in human nuclei. *J Cell Sci*. 1996;109:1427-1436.
- Pombo A, Cuello P, Schul W, et al. Regional and temporal specialization in the nucleus: a transcriptionally active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. *EMBO J*. 1998;17:1768-1778.
- Pombo A, Jackson DA, Hollinshead M, Wang Z, Roeder RG, Cook PR. Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J*. 1999;18:2241-2253.
- Grande MA, van der Kraan I, de Jong L, van Driel R. Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II. *J Cell Sci*. 1997;110:1781-1791.
- Strouboulis J, Wolffe AP. Functional compartmentalization of the nucleus. *J Cell Sci*. 1996;109:1991-2000.
- Neugebauer KM, Roth MB. Distribution of pre-mRNA splicing factors at sites of RNA polymerase II transcription. *Genes Dev*. 1997;11:1148-1159.
- Misteli T, Caceres JF, Spector DL. The dynamics of pre-mRNA splicing factor in living cells. *Nature*. 1997;387:523-527.
- Zeng C, Kim E, Warren SL, Berget SM. Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity. *EMBO J*. 1997;16:1401-1412.
- Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. *Science*. 1998;280:585-590.
- Kraus WL, Kadonaga JT. p300 and estrogen receptor cooperatively activate transcription via differential enhancement of initiation and reinitiation. *Genes Dev*. 1998;12:331-342.

14. Ho SN, Biggar SR, Spencer DM, Schreiber SL, Crabtree GR. Dimeric ligands define a role for transcriptional activation domains in reinitiation. *Nature*. 1996;382:822-826.
15. Sandaltzopoulos R, Becker PB. Heat shock factor increases the reinitiation rate from potentiated chromatin templates. *Mol Cell Biol*. 1998;18:361-367.
16. White J, Brou C, Wu J, Lutz Y, Moncollin V, Chambon P. The acidic transcriptional activator GAL-VP16 acts on preformed template-committed complexes. *EMBO J*. 1992;11:2229-2240.
17. Ross IL, Browne CM, Hume DA. Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol Cell Biol*. 1994;72:177-185.
18. Fiering S, Northrop JP, Nolan GP, Mattila PS, Crabtree GR, Herzenberg LA. Single cell assay of a transcription factor reveals a threshold in transcription activated by signals emanating from the T-cell antigen receptor. *Genes Dev*. 1990;4:1823-1834.
19. Ko MS, Nakauchi H, Takahashi N. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO J*. 1990;9:2835-2842.
20. Newlands S, Levitt LK, Robinson CS, et al. Transcription occurs in pulses in muscle fibers. *Genes Dev*. 1998;12:2748-2758.
21. White MR, Masuko M, Amet L, et al. Real-time analysis of the transcriptional regulation of HIV and hCMV promoters in single mammalian cells. *J Cell Sci*. 1995;108:441-455.
22. Takasuka N, White MR, Wood CD, Robertson WR, Davis JR. Dynamic changes in prolactin promoter activation in individual living lactotrophic cells. *Endocrinology*. 1998;139:1361-1368.
23. Vermaak D, Wolffe AP. Chromatin and chromosomal controls in development. *Dev Genet*. 1998;22:1-6.
24. Martin DI, Whitelaw E. The vagaries of variegating transgenes. *Bioessays*. 1996;18:919-923.
25. Robertson G, Garrick D, Wu W, Kearns M, Martin D, Whitelaw E. Position-dependent variegation of globin transgene expression in mice. *Proc Natl Acad Sci U S A*. 1995;92:5371-5375.
26. Sutherland HG, Martin DI, Whitelaw E. A globin enhancer acts by increasing the proportion of erythrocytes expressing a linked transgene. *Mol Cell Biol*. 1997;17:1607-1614.
27. Edmondson DG, Roth SY. Chromatin and transcription. *FASEB J*. 1996;10:1173-1182.
28. Paranjape SM, Kamakaka RT, Kadonaga JT. Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Annu Rev Biochem*. 1994;63:265-297.
29. Wolffe AP. The transcription of chromatin templates. *Curr Opin Genet Dev*. 1994;4:245-254.
30. Wolffe AP, Hayes JJ. Chromatin disruption and modification. *Nucleic Acids Res*. 1999;27:711-720.
31. Chi T, Lieberman P, Ellwood K, Carey M. A general mechanism for transcriptional synergy by eukaryotic activators. *Nature*. 1995;377:254-257.
32. Heller H, Bengal E. TFIID (TBP) stabilizes the binding of MyoD to its DNA site at the promoter and MyoD facilitates the association of TFIIB with the preinitiation complex. *Nucleic Acids Res*. 1998;26:2112-2119.
33. Hoopes BC, LeBlanc JF, Hawley DK. Contributions of the TATA box sequence to rate-limiting steps in transcription initiation by RNA polymerase II. *J Mol Biol*. 1998;277:1015-1031.
34. Huh JR, Park JM, Kim M, Carlson BA, Hatfield DL, Lee BJ. Recruitment of TBP or TFIIB to a promoter proximal position leads to stimulation of RNA polymerase II transcription without activator proteins both in vivo and in vitro. *Biochem Biophys Res Commun*. 1999;256:45-51.
35. Jenster G, Spencer TE, Burcin MM, Tsai SY, Tsai MJ, O'Malley BW. Steroid receptor induction of gene transcription: a two-step model. *Proc Natl Acad Sci U S A*. 1997;94:7879-7884.
36. Kim MH, Peterson DO. Oct-1 protein promotes functional transcription complex assembly on the mouse mammary tumor virus promoter. *J Biol Chem*. 1995;270:27823-27828.
37. Kim TK, Kim TH, Maniatis T. Efficient recruitment of TFIIB and CBP-RNA polymerase II holoenzyme by an interferon-beta enhanceosome in vitro. *Proc Natl Acad Sci U S A*. 1998;95:12191-12196.
38. Klages N, Strubin M. Stimulation of RNA polymerase II transcription initiation by recruitment of TBP in vivo. *Nature*. 1995;374:822-823.
39. Naar AM, Beaurang PA, Robinson KM, et al. Chromatin, TAFs, and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP-1 in vitro. *Genes Dev*. 1998;12:3020-3031.
40. Oelgeschlaeger T, Tao Y, Kang YK, Roeder RG. Transcription activation via enhanced preinitiation complex assembly in a human cell-free system lacking TAFs. *Mol Cell*. 1998;1:925-931.
41. Ranish JA, Yudkovsky N, Hahn S. Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes Dev*. 1999;13:49-63.
42. Walters MC, Fiering S, Eidemiller J, Magis W, Groudine M, Martin DI. Enhancers increase the probability but not the level of gene expression. *Proc Natl Acad Sci U S A*. 1995;92:7125-7129.
43. Barton MC, Madani N, Emerson BM. Distal enhancer regulation by promoter derepression in topologically constrained DNA in vitro. *Proc Natl Acad Sci U S A*. 1997;94:7257-7262.
44. Rothenberg EV, Ward SB. A dynamic assembly of diverse transcription factors integrates activation and cell-type information for interleukin 2 gene regulation. *Proc Natl Acad Sci U S A*. 1996;93:9358-9365.
45. Adams CC, Workman JL. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol Cell Biol*. 1995;15:1405-1421.
46. Ellwood K, Huang W, Johnson R, Carey M. Multiple layers of cooperativity regulate enhanceosome-responsive RNA polymerase II transcription complex assembly. *Mol Cell Biol*. 1999;19:2613-2623.
47. Kringstein AM, Rossi FM, Hofmann A, Blau HM. Graded transcriptional response to different concentrations of a single transactivator. *Proc Natl Acad Sci U S A*. 1998;95:13670-13675.
48. Ogawa M. Differentiation and proliferation of hematopoietic stem cells. *Blood*. 1993;81:2844-2853.
49. Ogawa M. Hematopoiesis. *J Allergy Clin Immunol*. 1994;94:645-650.
50. Nishijima I, Nakahata T, Watanabe S, et al. Hematopoietic and lymphopoietic responses in human granulocyte-macrophage colony-stimulating factor (GM-CSF) receptor transgenic mice injected with human GM-CSF. *Blood*. 1997;90:1031-1038.
51. Brady G, Billia F, Knox J, et al. Analysis of gene expression in a complex differentiation hierarchy by global amplification of cDNA from single cells. *Curr Biol*. 1995;5:909-922.
52. Hu M, Krause D, Greaves M, et al. Multilineage gene expression precedes commitment in the hematopoietic system. *Genes Dev*. 1997;11:774-785.
53. McKinstry WJ, Li CL, Rasko JE, Nicola NA, Johnson GR, Metcalf D. Cytokine receptor expression on hematopoietic stem and progenitor cells. *Blood*. 1997;89:65-71.
54. Xu D, Chan WL, Leung BP, et al. Selective expression of a stable cell surface molecule on type 2 but not type 1 helper T cells. *J Exp Med*. 1998;187:787-794.
55. Kelso A, Groves P, Trout AB, Francis K. Evidence for the stochastic acquisition of cytokine profile by CD4+ T cells activated in a T helper type 2-like response in vivo. *Eur J Immunol*. 1995;25:1168-1175.
56. Kelso A, Groves P, Ramm L, Doyle AG. Single-cell analysis by RT-PCR reveals differential expression of multiple type 1 and 2 cytokine genes among cells within polarized CD4+ T cell populations. *Int Immunol*. 1999;11:617-621.
57. Bucy RP, Karr L, Huang GQ, et al. Single cell analysis of cytokine gene expression coexpression during CD4+ T-cell phenotype development. *Proc Natl Acad Sci U S A*. 1995;92:7565-7569.
58. Openshaw P, Murphy EE, Hosken NA, et al. Heterogeneity of intracellular cytokine synthesis at the single cell level in polarized T helper 1 and T helper 2 populations. *J Exp Med*. 1995;182:1357-1367.
59. Saparov A, Elson CO, Devore-Carter D, Bucy RP, Weaver CT. Single-cell analyses of CD4+ T cells from alpha beta T cell receptor-transgenic mice: a distinct mucosal cytokine phenotype in the absence of transgene-specific antigen. *Eur J Immunol*. 1997;27:1774-1781.
60. Weaver CT, Saparov A, Kraus LA, Rogers WO, Hockett RD, Bucy RP. Heterogeneity in the clonal T cell response: implications for models of T cell activation and cytokine phenotype development. *Immunol Res*. 1998;17:279-302.
61. Rogers WO, Weaver CT, Kraus LA, Li J, Li L, Bucy RP. Visualization of antigen-specific T cell activation and cytokine expression in vivo. *J Immunol*. 1997;158:649-657.
62. Bucy RP, Panoskaltis-Mortari A, Huang GQ, et al. Heterogeneity of single cell cytokine gene expression in clonal T cell populations. *J Exp Med*. 1994;180:1251-1262.
63. Hollander GA, Zuklys S, Morel C, et al. Monoallelic expression of the interleukin-2 locus. *Science*. 1998;279:2118-2121.
64. Riviere I, Sunshine MJ, Littman DR. Regulation of IL-4 expression by activation of individual alleles. *Immunity*. 1998;9:217-228.
65. Naramura M, Hu RJ, Gu H. Mice with a fluorescent marker for interleukin 2 gene activation. *Immunity*. 1998;9:209-216.
66. Bix M, Locksley RM. Independent and epigenetic regulation of the interleukin-4 alleles in CD4+ T cells. *Science*. 1998;281:1352-1354.
67. Cook DL, Gerber AN, Tapscott SJ. Modeling stochastic gene expression: implications for haploinsufficiency. *Proc Natl Acad Sci U S A*. 1998;95:15641-15646.
68. Nutt SL, Vambrie S, Steinlein P, et al. Independent regulation of the two Pax5 alleles during B-cell development. *Nat Genet*. 1999;21:390-395.
69. McAdams HH, Arkin A. It's a noisy business! genetic regulation at the nanomolar scale. *Trends Genet*. 1999;15:65-69.
70. Sweet MJ, Hume DA. Endotoxin signal transduction in macrophages. *J Leukoc Biol*. 1996;60:8-26.
71. Fowles LF, Martin ML, Nelsen L, et al. Persistent activation of mitogen-activated protein kinases p42 and p44 and ets-2 phosphorylation in response to colony-stimulating factor 1/c-fms signaling. *Mol Cell Biol*. 1998;18:5148-5156.
72. Stacey KJ, Fowles LF, Colman MS, Ostrowski MC, Hume DA. Regulation of urokinase-type plasminogen activator gene transcription by macrophage colony-stimulating factor. *Mol Cell Biol*. 1995;15:3430-3441.
73. Hartmann G, Krieg AM. CpG DNA and LPS induce distinct patterns of activation in human monocytes. *Gene Ther*. 1999;6:893-903.
74. Sester DP, Beasley SJ, Sweet MJ, et al. Immunostimulatory DNA down-modulates CSF-1 receptor surface expression on murine macrophages with concomitant growth arrest and factor-independent survival. *J Immunol*. 1999;163:6541-6550.
75. Watt FM, Hogan BLM. Out of Eden: stem cells and their niches. *Science*. 2000;287:1427-1430.