Short residence times of DNA-bound transcription factors can reduce gene expression noise and increase the transmission of information in a gene regulation system

Eugenio Azpeitia[1,2,3], Andreas Wagner[1,2,4]

1. Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland.
2. Swiss Institute of Bioinformatics, Lausanne, Switzerland.
3. Centro de Ciencias Matemáticas, UNAM, Morelia, México.
4. The Santa Fe Institute, Santa Fe, NM, USA.

Abstract

Gene expression noise is not just ubiquitous but also variable, and we still do not understand some of the most elementary factors that affect it. Among them is the residence time of a transcription factor (TF) on DNA, the mean time that a DNA-bound TF remains bound. Here, we use a stochastic model of transcriptional regulation to study how this residence time affects gene expression. We find that the effect of residence time on gene expression depends on the level of induction of the gene. At high levels of induction, residence time has no effect on gene expression. However, as the level of induction decreases, short residence times reduce gene expression noise. The reason is that fast on-off TF binding dynamics prevent long periods where proteins are predominantly synthesized or degraded, which can cause excessive fluctuations in gene expression. As a consequence, short residence times can help a gene regulation system acquire information about the cellular environment it operates in. Our predictions are consistent with the observation that experimentally measured residence times are usually modest and lie between seconds to minutes.

Introduction

All gene expression is noisy. It produces mRNA and protein molecules whose numbers fluctuate randomly. Such noise is caused by stochastic molecular interactions, which include interactions between transcription factors and DNA, and by the stochastic synthesis and degradation of molecules (1,2). Gene expression noise affects multiple biological processes. For example, it can promote phenotypic diversity, influence the coordination of gene expression, trigger cell differentiation, and facilitate evolutionary transitions (1–7). Furthermore, noise can also reduce a cell's ability to acquire information about its environment. Such information is essential whenever cells need to respond to changing environments (8,9). It is acquired by signaling pathways that modulate the activity or concentration of transcription factors (TFs), which up-regulate or down-regulate effector genes. Thus, reducing gene expression noise can increase the ability of a regulated gene to capture information about a TF's changing concentration or activity, which is fundamental to produce an optimal cellular response to environmental change (10).

Gene expression regulation is being studied by many researchers whose insights improve our capacity to control and modify living systems (11–14). However, we still do not fully understand how some elementary properties of the interaction of a TF with its

1

binding site on DNA affect the stochastic dynamics of gene expression and the acquisition of information (15). One of these properties is a TF's residence time on its DNA binding site – the mean time that a TF remains bound to DNA. The residence time is equal to the inverse of the dissociation rate k_d between a TF and DNA. Theoretical and experimental work has shown that the dissociation rate can affect gene expression, affect the size of gene expression bursts (16–19), and modulate gene expression noise (20–23). However, it is difficult to discern whether the dissociation rate affects gene expression by altering the residence time or the affinity between a TF and DNA, because both depend on the dissociation constant k_d (affinity is given by the ratio $K_{eq}=k_d/k_a$ [M], where k_a is association rate between a TF and DNA).

Many TFs bind DNA transiently, with residence times ranging between seconds and minutes (24–35). Such TFs include MYC, p53, and glucocorticoid receptors, which are involved in fundamental processes such as apoptosis, DNA repair, DNA maintenance, and stress responses. They also include pioneer TFs that directly interact with chromatin and open it (36–38). The duration of a TF's residence time on a specific binding site can vary, even for different tissues within the same organisms, by chromatin modifications and the interaction of the TF with other molecular components (24,39,40). Such variation implies that residence time may play a role in regulating gene expression. Nevertheless, we do not know how residence time affects gene expression noise, because of limitations in experimental technology. For example, it is difficult to measure residence time and simultaneously quantify the rate of gene expression. It is also hard to quantify the number of TF binding events at a specific binding site in a given time (16,24,31). Moreover, it is challenging to experimentally modify only residence time without also affecting affinity.

Here we circumvent these limitations through a stochastic model of gene expression regulation. With this model, we study how residence time affects gene expression noise and the amount of information acquired by a gene expression system. Our analyses show that the effect of residence time increases as the level of induction of a gene decreases. At high induction levels, residence time has no effect on gene expression. However, as induction levels decrease, shorter residence times reduce the amount of gene expression noise and produce more regular gene expression dynamics. Shorter residence times also increase a gene regulation system's capacity to acquire information about the concentration of a TF. In sum, shorter residence times improve a gene's response to changes in its cellular environment.

Results

Model and main concepts

We use a two-state model of gene expression that represents the transcriptional activation and inactivation of a gene. In this model, TF molecules associate and dissociate from the gene's transcription factor binding site (DNA_{bs}) at rates k_a (M^{-1}s^{-1}) and k_d (s^{-1}), respectively. The regulated gene is expressed only when its binding site is bound by a TF, in which case the gene is transcribed into mRNA at rate k_1. The resulting mRNA is then translated into protein molecules at rate k_2. Finally, mRNA and protein molecules degrade at rates d_1 and d_2, respectively (Fig. 1a).

Residence time is the average life span or half-life ($t_{1/2}$) of the TF-DNA_{bs} complex. In other words, residence time quantifies the stability of this complex (Fig. 1b). Affinity is quantified with the equilibrium constant K_{eq}. The equilibrium constant is equal to the concentration of free TF at which half of all binding sites are occupied. A high equilibrium constant is equal to a low affinity, because it means that a large concentration of TF is required to occupy 50% of binding sites. For a given TF concentration, the probability that a binding site is occupied increases with increasing affinity (Fig. 1c).

Although both affinity and residence time depend on the dissociation rate k_d, they can be modified independently from each other. Changing the dissociation rate will modify the residence time, but it can leave the affinity unchanged if the association rate changes appropriately to keep the ratio k_d/k_a constant. Conversely, by changing only the association rate, affinity can be modified without altering residence time.

Notice that affinity and TF concentration jointly determine the level of induction of gene expression, because a gene is more likely to be active when the concentration of free TF is higher than its affinity to DNA. Conversely, when this concentration is lower than the affinity, the gene will tend to be inactive. One can increase the level of a gene's induction by increasing either TF concentration or affinity (Fig. 1d). Below we change the level of induction by modifying TF concentration, but changing affinity itself yields the same observations (see Sup Inf 1).

We simulate gene regulation dynamics using Gillespie's algorithm (41), which reproduces the stochastic dynamics of many chemical systems, using biologically meaningful values of all biochemical parameters (Sup Table 1 and Methods). Because both the mRNA and protein output of our modeled gene regulation system behave qualitatively identically, we focus on the protein output below (see Sup Figs for mRNA).

Short residence times reduce gene expression noise and modulate gene expression dynamic

We first study the effect of residence time and affinity on gene expression noise. To quantify noise, we quantified the size of the temporal fluctuations in the number of proteins, as the difference between the maximal and the minimal number of expressed protein molecules ($N_{Pmax} - N_{Pmin}$), and averaged this difference over 1000 simulations. Two alternative noise measures, the coefficient of variation and the Fano factor yield identical observations (see Sup Inf 2 and Sup Fig 1).

In these simulations, we varied residence time within the interval [1s,1h], TF concentration within the interval [10^{-11}M,10^{-7}M], and set the affinity to 10^{-9}M. Notice that the TF concentration interval ranges two orders of magnitude below and above the affinity, which implies that the level of gene induction ranges from almost always inactive to almost always active. Hence, high and low TF concentration values correspond to high and low induction levels, respectively.

At the highest TF concentration, residence time does not affect noise (Fig. 2a; Sup Fig. 2a). However, as the TF concentration decreases, a longer residence time increases noise (Fig. 2a; Sup Fig. 2a). For example, individual protein expression trajectories at

extremely short ($t_{1/2}$=1s) and long ($t_{1/2}$=1h) residence times are very similar at the highest TF concentration (Fig. 2b; Sup Fig. 2b). However, as TF concentration decreases, long residence times cause sporadic, but large fluctuations in protein expression (Fig. 2c and d; Sup Fig. 2c and d).

To understand these observations, notice that when the TF concentration is very high, the level of induction is high because a TF molecule is bound to DNA most of the time. In this case, gene expression resembles that of a constitutive gene, regardless of the TF's residence time. In consequence, gene expression noise is only determined by the degradation and synthesis rates of mRNA and protein (1,2). In other words, it is independent of residence time (Fig. 2a and b; Sup Fig. 2a and d).

This is no longer true as the TF concentration decreases. In this case, the probability that a TF is bound to DNA at any one time decreases, and longer residence times increase the average amount of time that a TF is either bound or unbound. In other words, longer residence times produce longer periods of active and inactive gene expression. During active periods, proteins are produced, whereas during inactive periods, previously expressed proteins decay. Thus, longer residence times lengthen both active and inactive periods, which results in large fluctuations in the number of proteins (Fig. 2c and d; Sup Fig. 2c and d). Reducing the residence time (at constant induction) decreases the duration of both active and inactive periods by the same amounts. As a result, expressed molecules accumulate and decay for shorter time periods, and fluctuations in these molecules become smaller (Fig. 2a-d; Sup Fig. 2a-d).

In contrast to its effects on noise, residence time does not affect the mean level of protein expression, which only depends on the level of induction (Fig. 2e; Sup Fig. 2e). The reason is that the frequency of both protein production and degradation events (i.e., the mean number of proteins produced and degraded in a given period of time) is not affected by residence time, regardless of the level of induction (Fig. 2f; Sup Fig. 2f and 3a and b). However, as the level of induction decreases, shorter residence times decrease variation in the frequency of production and degradation events (Fig. 2f; Sup Fig. 2f and 3a and b). As a consequence, protein expression is more homogeneous at shorter residence times, because protein production and degradation events alternate more regularly, except at the very lowest affinities (Sup Inf 3 and Sup Fig. 3c-f).

Residence time and information

Due to noise, the regulation of gene expression transforms a concentration of a TF into a distribution of expressed mRNA and protein molecules. Different TF concentrations may produce overlapping distributions of expressed molecules, in which case information about TF concentration gets lost. Reducing gene expression noise can reduce this overlap and thus also the amount of lost information. (Fig. 3a) (10). Because shorter residence times reduce gene expression noise (Fig. 2), we hypothesized that they also increase the amount of information protein which expression levels contain about the concentration of the regulating TF. To find out, we quantified the mutual information between protein expression and TF concentration. Mutual information is an information theoretical quantity that encapsulates the reduction in uncertainty about one random variable provided by knowledge about another random variable (42) (see Methods). To quantify mutual information, we performed 2500 stochastic simulations of gene

4

expression dynamics for each of n evenly distributed TF concentrations within the interval $[10^{-7}M/n, 10^{-7}M]$, exploring affinity values within the interval $[10^{-12}M, 10^{-4}M]$. This range includes affinity values below the minimal TF concentration, where induction is low regardless of TF concentration, and above the maximal concentration, where induction is high regardless of TF concentration.

In earlier work, we have shown that the amount of information that gene expression levels contain about the concentration of the regulating TF depends on the affinity between a TF and its binding site. At very low affinities, gene regulation is insensitive to TF concentration, such that gene expression conveys little information. At very high affinities, the level of gene induction is high for all TF concentrations, such that gene expression also conveys little information – it is similar to that of a constitutively expressed gene for all TF concentrations (43). Our current work shows that this pattern holds regardless of residence time (Fig. 3b; Sup. Fig. 4a).

In contrast, residence time does affect acquired information at intermediate affinities. Specifically, even though the mean number of expressed molecules does not depend on a TF's residence time on DNA (Fig. 2e; Sup. Fig. 2e), their variability decreases with shorter residence times (Fig. 2a). As a result, as residence time decreases, the overlap between protein distributions decrease (Fig. 3c and d; Sup Fig. 4b and c), which increases the amount of acquired information (Fig. 3b; Sup Fig. 4a). In sum, under conditions where a gene regulation system can acquire information, shortening residence time increases the amount of acquired information.

Discussion

Previous theoretical and experimental work showed that gene expression noise can be modulated by the dissociation rate k_d of a DNA-bound TF (16–23), but this work did not distinguish between the effects of residence time and affinity. This distinction, however, is important because both properties depend on the dissociation rate but have different effects on gene expression dynamics. Here, we study these properties separately, and show that short residence time can reduce expression noise, while in general noise increases as affinity decreases (Fig 2).

Our results also show that the effect of residence time on gene expression is not independent from affinity. When a gene is highly induced, residence time does not affect gene expression. However, as the level of induction decreases, short residence times can help produce less noisy expression. Short residence times effectively fragment gene activity into short periods of active and inactive expression, which prevents the excessive accumulation and depletion of proteins, and thus also excessive stochastic variation in gene expression. Similarly, the effect of affinity on noise depends on residence time. In particular, the effect of affinity is reduced when residence time decreases. However, noise can even decrease with decreasing affinity when residence times are very short and in the range of seconds (Fig. 2a).

In previous work, we showed that expressed proteins harbor information about the concentration of a TF regulating their expression, if the TF's affinity to regulatory DNA is of the same order of magnitude as its concentration (43). Here, we show that this behavior is independent of residence time. However, because shorter residence times

render gene expression less noisy, the overlap between protein expression distributions resulting from different TF concentrations decreases. Consequently, shorter residence times increase the ability of a gene regulation system to distinguish between different TF concentrations.

Recent technological advances will permit experimental verification of our observations. First, in the last decade methods to quantify the ability of a regulated gene to acquire information about the concentration of a TF have become standardized (10,44). Also, the effect of residence time on noise can be quantified with techniques that simultaneously quantify TF-DNA binding events and the production of mRNA molecules, which are being developed (16). It is especially difficult to obtain multiple alleles of a TF with different residence times without also altering the TF's affinity to DNA. However, because our model shows that affinity affects gene expression by changing the induction level at a specific TF concentration, our observations hold regardless of whether one varies TF concentration or affinity (Sup Inf 1). Hence, to test the effect of residence time on noise and information acquisition, one can compensate for any change in affinity by adjusting a gene's induction level with TF concentration.

Our two state model of gene regulation is simple and does not represent processes such as the binding of RNA polymerases to DNA explicitly (45). However, previous work has shown that the two state model produces similar gene expression dynamics as more complex models (19), suggesting that our main results may hold for such models. Nevertheless, each additional regulatory requires time, which may constrain the amount of time that a TF must be minimally bound to DNA before it can affect gene expression (46). Future work also needs to consider other kinetic parameters regulating gene expression, such as the mRNA synthesis rate, because noise and information can be affected by these parameters (1,2,20,22,47–49).

Our results are in agreement with previous work and complement this work. For example, experimental evidence suggests that the affinity of essential regulators of gene expression, such as NF-κB and TBP, modulates gene expression noise (50,51). Moreover, a model based on the binding dynamics of Sox2 and Oct4, two important regulators of the pluripotency of stems cells, showed that long residence times reduce the sensitivity of gene expression to TF concentration, because TFs with long residence times are bound to DNA most of the time regardless of their concentration (26,31). Another study showed that negative regulatory feedback loops in general suppress noise more effectively when residence times are short (23).

Our work also helps solve an apparent contradiction between experimental and theoretical work about the importance of residence time. In particular, it has been predicted that longer residence times facilitate gene expression, because they increase the probability of a successful activation of gene expression by TFs, by providing longer time for other components, such as polymerases, to successfully bind DNA (24,25,52–54). Experimentally measured residence times, which are generally short and lie within seconds to minutes (24–35), are inconsistent with this prediction. Our work shows that residence time does not affect average gene expression levels. Instead, residence time reduces expression noise and can help signaling systems acquire information without modifying the probability of a successful activation of gene expression, which can help explain why short residence times may be prevalent in nature.

Methods

Two-state model of gene expression

To study how a transcription factor's (TF's) residence time on DNA affects gene expression, we built a gene expression model in which a *TF* binds to a DNA binding site (DNA_{bs}) to regulate the expression of a nearby gene. *TF* molecules associate with the DNA binding site at a rate k_a ($M^{-1}s^{-1}$), and dissociate from it at a rate k_d (s^{-1}). Only when the TF is bound to DNA does transcription occur (at a rate k_1 [s^{-1}]). Transcribed mRNA molecules are degraded at a rate d_1 (s^{-1}). Proteins are translated from mRNA molecules at a rate k_2 (s^{-1}), and degraded at a rate d_2 (s^{-1}).

Stochastic simulations

To simulate the behavior of our gene expression model, we use Gillespie's discrete stochastic simulation algorithm (41), using the numpy python package for scientific computing (http://www.numpy.org/). Gillespie's algorithm captures the stochastic nature of chemical systems. It assumes a well-stirred and thermally equilibrated system with constant volume and temperature. The algorithm requires the probability p_j that a chemical reaction R_j occurs in a given time interval [t,t+τ). This probability p_j is proportional to both the reaction rate and the number of reacting molecules. For the reversible bindings of TF molecules to DNA, the association probability p_a and the dissociation probability p_d are given by

$$p_a = \frac{k_a}{V N_A} N_T N_D$$
$$p_d = k_d N_{TD},$$

where V is the reaction volume, N_A is Avogadro's number, and N_T, N_D, and N_{TD} are the numbers of TF molecules, DNA binding sites, and TF-DNA_{bs} complexes. Notice that the dissociation of TF-DNA_{bs} complexes is a first-order reaction, which is independent of the volume in which the reaction takes place. In contrast, the association of TF molecules with DNA binding sites is a second-order reaction, which is inversely proportional to the volume.

The probabilities p_{mRs}, p_{mRd}, p_{Ps} and p_{Pd} of mRNA transcription, mRNA degradation, protein synthesis, and protein degradation are given by

$$p_{mR_s} = k_1 N_{TD}$$
$$p_{mR_d} = d_1 N_{mR}$$
$$p_{P_s} = k_2 N_{mR}$$
$$p_{P_d} = d_2 N_P,$$

respectively. In these expressions, the quantities N_{TD}, N_{mR} and N_P are the numbers of TF-DNA complexes, mRNA molecules and of protein molecules, respectively. Because we model a haploid organism with only a single non-leaky DNA binding site, the probability of mRNA synthesis can be reduced to

7

$$p_{mR_s} = k_1$$

when the DNA is bound by the transcription factor (N_{TD}=1), and to

$$p_{mR_s} = 0$$

when it is unbound (N_{TD}=0).

Initial conditions for simulations

We assume that the TF a concentration of 10^{-7}M corresponds to a few thousand TF molecules per cell, a realistic number in animal and yeast cells (55,56). Because we model only one binding site, the concentration of free TF is not substantially affected by the binding of a single TF molecule to *DNA*. We therefore do not distinguish between the free and the total TF concentration. After this simplification, to determine the initial conditions of our model, we calculate the probability $\widehat{N}_{TD}$ that the binding site is bound by a TF molecule as

$$\widehat{N}_{TD} = \frac{N_T}{K_{eq} + N_T}$$

where N_T is the total number of TF molecules. We selected the initial state of the DNA (N_{TD_i}) at random with binomial probability $\widehat{N}_{TD}$ (i.e., $N_{TD_i} = 1$ if the DNA is bound by a TF molecule, and zero otherwise. It follows that

$$N_{D_i} = 1 - N_{TD_i}$$
$$N_{T_i} = N_T - N_{TD_i}$$

where N_{D_i} is the initial state of the number of non-bound DNA binding site and N_{T_i} is the number of free TF molecules. As the initial state of the number mRNA and protein molecules we used

$$N_{mR_i} = \widehat{N}_{TD} \frac{k_1}{d_1}$$
$$N_{P_i} = \widehat{N}_{TD} \frac{k_1}{d_1} \frac{k_2}{d_2},$$

which are the expected average number of mRNA and protein molecules for a constitutively expressed gene (2), multiplied by the probability that the DNA is bound by a TF molecule.

Information quantification

The number of molecules of any chemical species in a cell or in a unit volume fluctuates, because molecules are produced and decay stochastically. We use Shannon's entropy to quantify the unpredictability caused by such stochastic fluctuations in the number of transcription factor molecules as

$$H(\text{Pr}(TF)) = -\sum_{N_T = \frac{N_{Tmax}}{n}}^{N_{Tmax}} p(N_T)\log_2 p(N_T),$$

where $\text{Pr}(TF)$ is the distribution of the number of transcription factor molecules (N_T), and $p(N_T)$ is the probability that the system contains N molecules of the transcription factor.

To estimate information we performed simulations from n different numbers of transcription factor molecules that were evenly distributed within the interval $[N_{Tmax}/n, N_{Tmax}]$ ($n < N_{Tmax}$). For this reason

$$H(\text{Pr}(TF)) = \log_2 n.$$

TF-DNA binding triggers the transcription of mRNA molecules that are then translated into protein molecules. We use the number of mRNA N_{mRNA} and protein molecules N_P as the system's response or output, which we denote as O.

A gene expression system acquires information when the number of expressed proteins or mRNA reflects the number of transcription factors. This information can be quantified via the mutual information

$$I(TF; O) = H(\text{Pr}(TF)) - H(\text{Pr}(TF \mid O)),$$

a widely used quantity in information theory (42). It is equal to the difference between the entropy $H(\text{Pr}(TF))$ and the conditional entropy $H(\text{Pr}(TF|O))$, which represents the entropy of the number of transcription factor molecules for a given number of mRNA or protein molecules. In other words, the mutual information I quantifies the amount of information that the number of expressed mRNA or protein molecules harbors about the number of transcription factor molecules.

Noise quantification

The model produces a probability distribution of the number of mRNA and protein molecules for any given number of transcription factor molecules N_T. This response is thus a conditional probability distribution, which we write as

$$\text{Pr}(N_{O_{min}} \leq N_O \leq N_{O_{max}} | TF = N_T) \text{Pr}(N_{Omin} < N_O < N_{Omax} | TF = N_T),$$

where N_{Omin} and N_{Omax} are the minimal and maximal number of mRNA or protein molecules, respectively. We performed 1000 simulations to estimate noise using three different measures. The size of the fluctuation, Fano factor and the coefficient of variation. The size of the fluctuations was quantified as the average difference ($N_{O_{max}} - N_{O_{min}}$). Fano factor as the variance of the response distribution divided by its mean ($\sigma^2(N_P)/\bar{N}_P$). Coefficient of variation as the standard deviation of the response distribution divided by its mean ($\sigma(N_P)/\bar{N}_P$).

Parameter values

Our simulations considered biologically sensible parameter ranges. Specifically, for TF-DNA_{bs} binding, empirical data suggests that usually $K_{eq}<10^{-8}$ and can reach picomolar (10^{-12}M) or even smaller values (46,55,57–60). Because experimental research has shown that residence time varies from seconds to tens of minutes (24,27,29), we used dissociation rates producing residence time within this interval [1s,1h].

For mRNA, experimentally measured half-lives usually lie in the range of seconds to hours (61–64). Protein half-lives are usually longer than mRNA half-lives (65) and lie between hours and days (63,66). Taking all this information into consideration we chose mRNA half-lives of ~3.3min, and protein half-lives of ~1.5h. We assumed that the ratio k_2/k_1 of the protein synthesis rate to the mRNA synthesis rate exceeds 1.0 (67).

References

1.   Kepler TB, Elston TC. Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations. Biophysical Journal. 2001 Dec;81(6):3116–36.

2.   Raj A, van Oudenaarden A. Single-Molecule Approaches to Stochastic Gene Expression. Annual Review of Biophysics. 2009 Jun;38(1):255–70.

3.   Sanchez A, Choubey S, Kondev J. Regulation of Noise in Gene Expression. Annu Rev Biophys. 2013 May 6;42(1):469–91.

4.   Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. Nature. 2010 Sep;467(7312):167–73.

5.   Fraser D, Kaern M. A chance at survival: gene expression noise and phenotypic diversification strategies. Molecular Microbiology. 2009 Mar;71(6):1333–40.

6.   Urban EA, Johnston RJ. Buffering and Amplifying Transcriptional Noise During Cell Fate Specification. Front Genet. 2018 Nov 29;9:591.

7.   Karig D, Martini KM, Lu T, DeLateur NA, Goldenfeld N, Weiss R. Stochastic Turing patterns in a synthetic bacterial population. Proc Natl Acad Sci USA. 2018 Jun 26;115(26):6572–7.

8.   Bowsher CG, Swain PS. Environmental sensing, information transfer, and cellular decision-making. Current Opinion in Biotechnology. 2014 Aug;28:149–55.

9.   Levchenko A, Nemenman I. Cellular noise and information transmission. Current Opinion in Biotechnology. 2014 Aug;28:156–64.

10.   Rhee A, Cheong R, Levchenko A. The application of information theory to biochemical signaling systems. Physical Biology. 2012 Aug 1;9(4):045011.

11.   Pope SD, Medzhitov R. Emerging Principles of Gene Expression Programs and Their Regulation. Molecular Cell. 2018 Aug;71(3):389–97.

12.   Sainsbury S, Bernecky C, Cramer P. Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol. 2015 Mar;16(3):129–43.

13.   Lelli KM, Slattery M, Mann RS. Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. Annu Rev Genet. 2012 Dec 15;46(1):43–68.

14.   Li B, Carey M, Workman JL. The Role of Chromatin during Transcription. Cell. 2007 Feb;128(4):707–19.

15.   Einav T, Phillips R. How the avidity of polymerase binding to the –35/–10 promoter sites affects gene expression. Proc Natl Acad Sci USA. 2019 Jul 2;116(27):13340–5.

16.   Donovan BT, Huynh A, Ball DA, Patel HP, Poirier MG, Larson DR, et al. Live-cell

imaging reveals the interplay between transcription factors, nucleosomes, and bursting. EMBO J. 2019 Jun 17;38(12):e100809.

17.   Fujita K, Iwaki M, Yanagida T. Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA. Nat Commun. 2016 Dec;7(1):13788.

18.   Skupsky R, Burnett JC, Foley JE, Schaffer DV, Arkin AP. HIV Promoter Integration Site Primarily Modulates Transcriptional Burst Size Rather Than Frequency. Friedman N, editor. PLoS Computational Biology. 2010 Sep 30;6(9):e1000952.

19.   Kumar N, Singh A, Kulkarni RV. Transcriptional Bursting in Gene Expression: Analytical Results for General Stochastic Models. Morozov AV, editor. PLoS Comput Biol. 2015 Oct 16;11(10):e1004292.

20.   Raser JM. Control of Stochasticity in Eukaryotic Gene Expression. Science. 2004 Jun 18;304(5678):1811–4.

21.   Peccoud J, Ycart B. Markovian Modeling of Gene-Product Synthesis. Theoretical Population Biology. 1995 Oct;48(2):222–34.

22.   Iyer-Biswas S, Hayot F, Jayaprakash C. Stochasticity of gene products from transcriptional pulsing. Phys Rev E. 2009 Mar 23;79(3):031911.

23.   Grönlund A, Lötstedt P, Elf J. Transcription factor binding kinetics constrain noise suppression via negative feedback. Nat Commun. 2013 Oct;4(1):1864.

24.   Mueller F, Stasevich TJ, Mazza D, McNally JG. Quantifying transcription factor kinetics: At work or at play? Critical Reviews in Biochemistry and Molecular Biology. 2013 Sep;48(5):492–514.

25.   Clauß K, Popp AP, Schulze L, Hettich J, Reisser M, Escoter Torres L, et al. DNA residence time is a regulatory factor of transcription repression. Nucleic Acids Research. 2017 Nov 2;45(19):11121–30.

26.   Chen J, Zhang Z, Li L, Chen B-C, Revyakin A, Hajj B, et al. Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. Cell. 2014 Mar;156(6):1274–85.

27.   Morisaki T, Müller WG, Golob N, Mazza D, McNally JG. Single-molecule analysis of transcription factor binding at transcription sites in live cells. Nature Communications [Internet]. 2014 Dec [cited 2019 Aug 9];5(1). Available from: http://www.nature.com/articles/ncomms5456

28.   Mazza D, Abernathy A, Golob N, Morisaki T, McNally JG. A benchmark for chromatin binding measurements in live cells. Nucleic Acids Research. 2012 Aug 1;40(15):e119–e119.

29.   Geertz M, Shore D, Maerkl SJ. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. Proceedings of the National Academy of Sciences. 2012 Oct 9;109(41):16540–5.

30.   Gebhardt JCM, Suter DM, Roy R, Zhao ZW, Chapman AR, Basu S, et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. Nat Methods. 2013 May;10(5):421–6.

31.   Liu Z, Tjian R. Visualizing transcription factor dynamics in living cells. J Cell Biol. 2018 Apr 2;217(4):1181–91.

32.   Rullan M, Benzinger D, Schmidt GW, Milias-Argeitis A, Khammash M. An Optogenetic Platform for Real-Time, Single-Cell Interrogation of Stochastic Transcriptional Regulation. Molecular Cell. 2018 May;70(4):745-756.e6.

33.   Deluz C, Friman ET, Strebinger D, Benke A, Raccaud M, Callegari A, et al. A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. Genes Dev. 2016 Nov 15;30(22):2538–50.

34.   Sugo N, Morimatsu M, Arai Y, Kousoku Y, Ohkuni A, Nomura T, et al. Single-

Molecule Imaging Reveals Dynamics of CREB Transcription Factor Bound to Its Target Sequence. Sci Rep. 2015 Sep;5(1):10662.

35.   Hager G. Footprints by deep sequencing. Nat Methods. 2009 Apr;6(4):254–5.

36.   Joerger AC, Fersht AR. The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches. Annu Rev Biochem. 2016 Jun 2;85(1):375–404.

37.   Oakley RH, Cidlowski JA. The biology of the glucocorticoid receptor: New signaling mechanisms in health and disease. Journal of Allergy and Clinical Immunology. 2013 Nov;132(5):1033–44.

38.   Swinstead EE, Miranda TB, Paakinaho V, Baek S, Goldstein I, Hawkins M, et al. Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. Cell. 2016 Apr;165(3):593–605.

39.   Giglia-Mari G, Theil AF, Mari P-O, Mourgues S, Nonnekens J, Andrieux LO, et al. Differentiation Driven Changes in the Dynamic Organization of Basal Transcription Initiation. Misteli T, editor. PLoS Biol. 2009 Oct 20;7(10):e1000220.

40.   van Werven FJ, van Teeffelen HAAM, Holstege FCP, Timmers HTM. Distinct promoter dynamics of the basal transcription factor TBP across the yeast genome. Nat Struct Mol Biol. 2009 Oct;16(10):1043–8.

41.   Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry. 1977 Dec;81(25):2340–61.

42.   Cover TM, Thomas JA. Elements of information theory. 2nd ed. Hoboken, N.J: Wiley-Interscience; 2006. 748 p.

43.   Azpeitia E, Wagner A. The positive role of noise for information acquisition in biological signaling pathways [Internet]. Systems Biology; 2019 Sep [cited 2019 Sep 10]. Available from: http://biorxiv.org/lookup/doi/10.1101/762989

44.   Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A. Information Transduction Capacity of Noisy Biochemical Signaling Networks. Science. 2011 Oct 21;334(6054):354–8.

45.   Cramer P. Organization and regulation of gene transcription. Nature [Internet]. 2019 Aug 28 [cited 2019 Sep 2]; Available from: http://www.nature.com/articles/s41586-019-1517-4

46.   Corzo J. Time, the forgotten dimension of ligand binding teaching. Biochemistry and Molecular Biology Education. 2006 Nov;34(6):413–6.

47.   ten Wolde PR, Becker NB, Ouldridge TE, Mugler A. Fundamental Limits to Cellular Sensing. Journal of Statistical Physics. 2016 Mar;162(5):1395–424.

48.   Tkacik G, Callan CG, Bialek W. Information flow and optimization in transcriptional regulation. Proceedings of the National Academy of Sciences. 2008 Aug 26;105(34):12265–70.

49.   Rieckh G, Tkačik G. Noise and Information Transmission in Promoters with Multiple Internal States. Biophysical Journal. 2014 Mar;106(5):1194–204.

50.   Ravarani CNJ, Chalancon G, Breker M, de Groot NS, Babu MM. Affinity and competition for TBP are molecular determinants of gene expression noise. Nat Commun. 2016 Apr;7(1):10417.

51.   Heltberg ML, Krishna S, Jensen MH. On chaotic dynamics in transcription factors and the associated effects in differential gene regulation. Nat Commun. 2019 Dec;10(1):71.

52.   Lickwar CR, Mueller F, Hanlon SE, McNally JG, Lieb JD. Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function. Nature. 2012 Apr;484(7393):251–5.

53.   Stavreva DA, Muller WG, Hager GL, Smith CL, McNally JG. Rapid Glucocorticoid

Receptor Exchange at a Promoter Is Coupled to Transcription and Regulated by Chaperones and Proteasomes. Molecular and Cellular Biology. 2004 Apr 1;24(7):2682–97.

54.   Gorski SA, Snyder SK, John S, Grummt I, Misteli T. Modulation of RNA Polymerase Assembly Dynamics in Transcriptional Regulation. Molecular Cell. 2008 May;30(4):486–97.

55.   Biggin MD. Animal Transcription Networks as Highly Connected, Quantitative Continua. Developmental Cell. 2011 Oct;21(4):611–26.

56.   Ho B, Baryshnikova A, Brown GW. Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. Cell Systems. 2018 Feb;6(2):192-205.e3.

57.   Jung J-H, Lee H-J, Ryu JY, Park C-M. SPL3/4/5 Integrate Developmental Aging and Photoperiodic Signals into the FT-FD Module in Arabidopsis Flowering. Molecular Plant. 2016 Dec 5;9(12):1647–59.

58.   Belikov S, Berg OG, Wrange Ö. Quantification of transcription factor-DNA binding affinity in a living cell. Nucleic Acids Research. 2015 Dec 10;44(7):3045–58.

59.   Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmann R, et al. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. Proceedings of the National Academy of Sciences. 2012 Dec 26;109(52):21330–5.

60.   Li X-Y, Thomas S, Sabo PJ, Eisen MB, Stamatoyannopoulos JA, Biggin MD. The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. Genome Biology. 2011;12(4):R34.

61.   Lugowski A, Nicholson B, Rissland OS. Determining mRNA half-lives on a transcriptome-wide scale. Methods. 2018 Mar;137:90–8.

62.   Laalami S, Zig L, Putzer H. Initiation of mRNA decay in bacteria. Cellular and Molecular Life Sciences. 2014 May;71(10):1799–828.

63.   Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011 May;473(7347):337–42.

64.   Sharova LV, Sharov AA, Nedorezov T, Piao Y, Shaik N, Ko MSH. Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells. DNA Research. 2009 Jan 8;16(1):45–58.

65.   Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, Chan L-K, et al. Sizing up metatranscriptomics. ISME J. 2013 Feb;7(2):237–43.

66.   Toyama BH, Hetzer MW. Protein homeostasis: live long, won't prosper. Nature Reviews Molecular Cell Biology. 2013 Jan;14(1):55–61.

67.   Hausser J, Mayo A, Keren L, Alon U. Central dogma rates and the trade-off between precision and economy in gene expression. Nature Communications [Internet]. 2019 Dec [cited 2019 Aug 9];10(1). Available from: http://www.nature.com/articles/s41467-018-07391-8
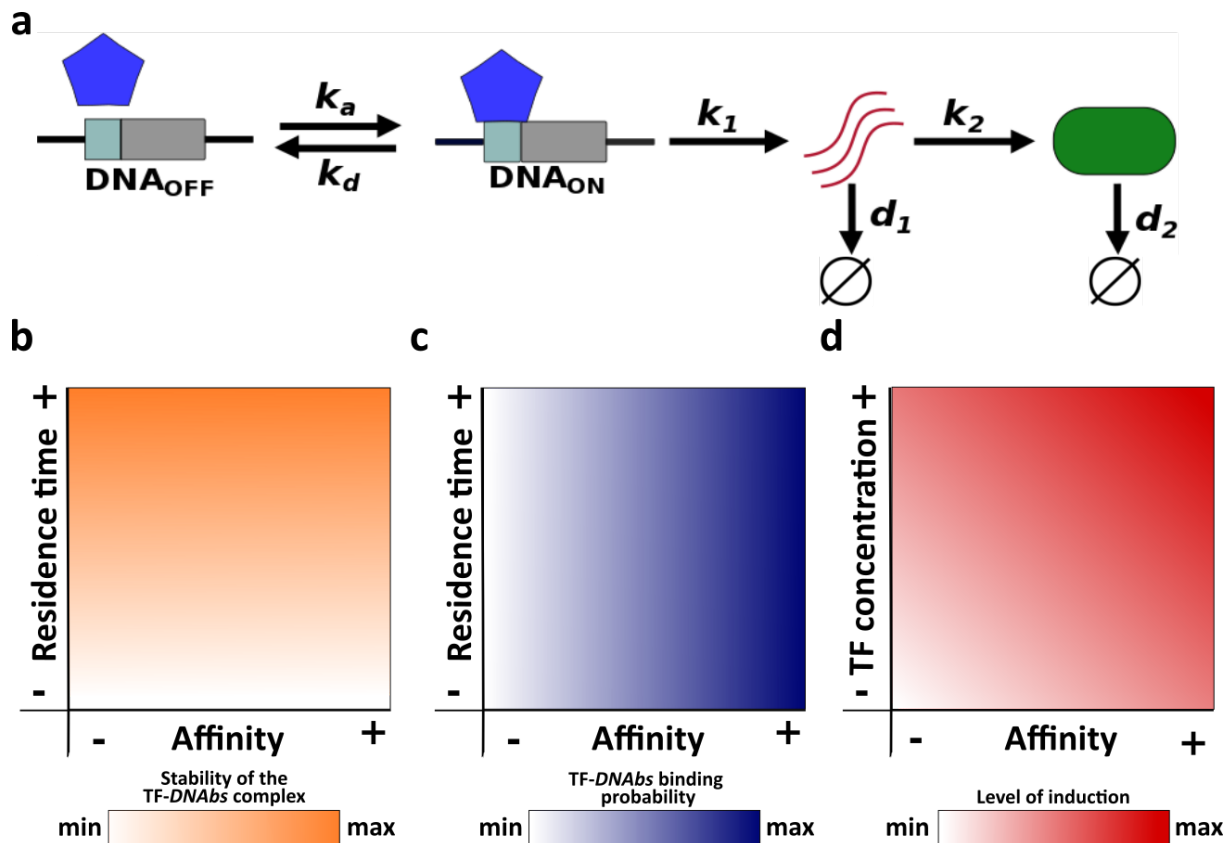
Figures

Fig 1. Schematic description of the model and main concepts. (a) k_a and k_d correspond to the association and dissociation rate, respectively; k_1 and k_2 correspond to the mRNA and protein synthesis rate, respectively; d_1 and d_2 correspond to the mRNA and protein degradation rate, respectively. Relationships of both residence time and affinity with (b) the stability of the TF-DNA_{bs} complex, and (c) TF-DNA_{bs} binding probability. (d) Relationship of affinity and TF concentration with the level of induction.
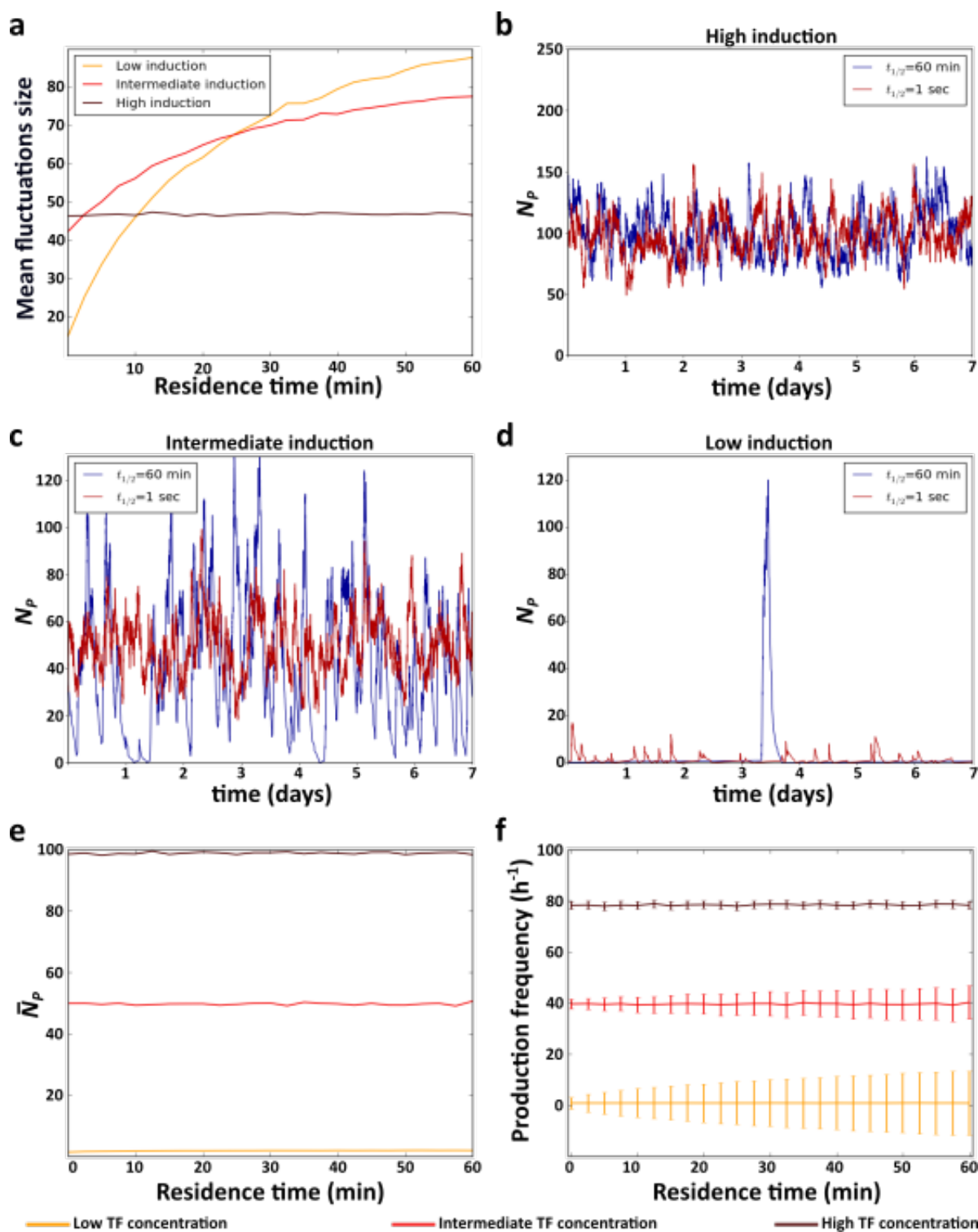
Fig 2. Effect of residence time on the size of the fluctuations in expressed protein molecules. (a) Mean fluctuation size in the number of protein molecules (y axis) at different levels of induction as a function of residence time (x axis). (b-d) Example time trajectories of the number of protein molecules N_P at three different levels of induction. Analyses of (e) mean number of protein ($\bar{N}_P$) (f) mean and coefficient of variation of the frequency of protein production events. (b-d) Red and blue lines show data for short (1s) and a long (1h) residence times, respectively. (a, e and f) High (TF=10^{-7}M), intermediate (TF=10^{-9}M) and low (TF=10^{-11}M) levels of induction are indicated in the color legend below the figure.
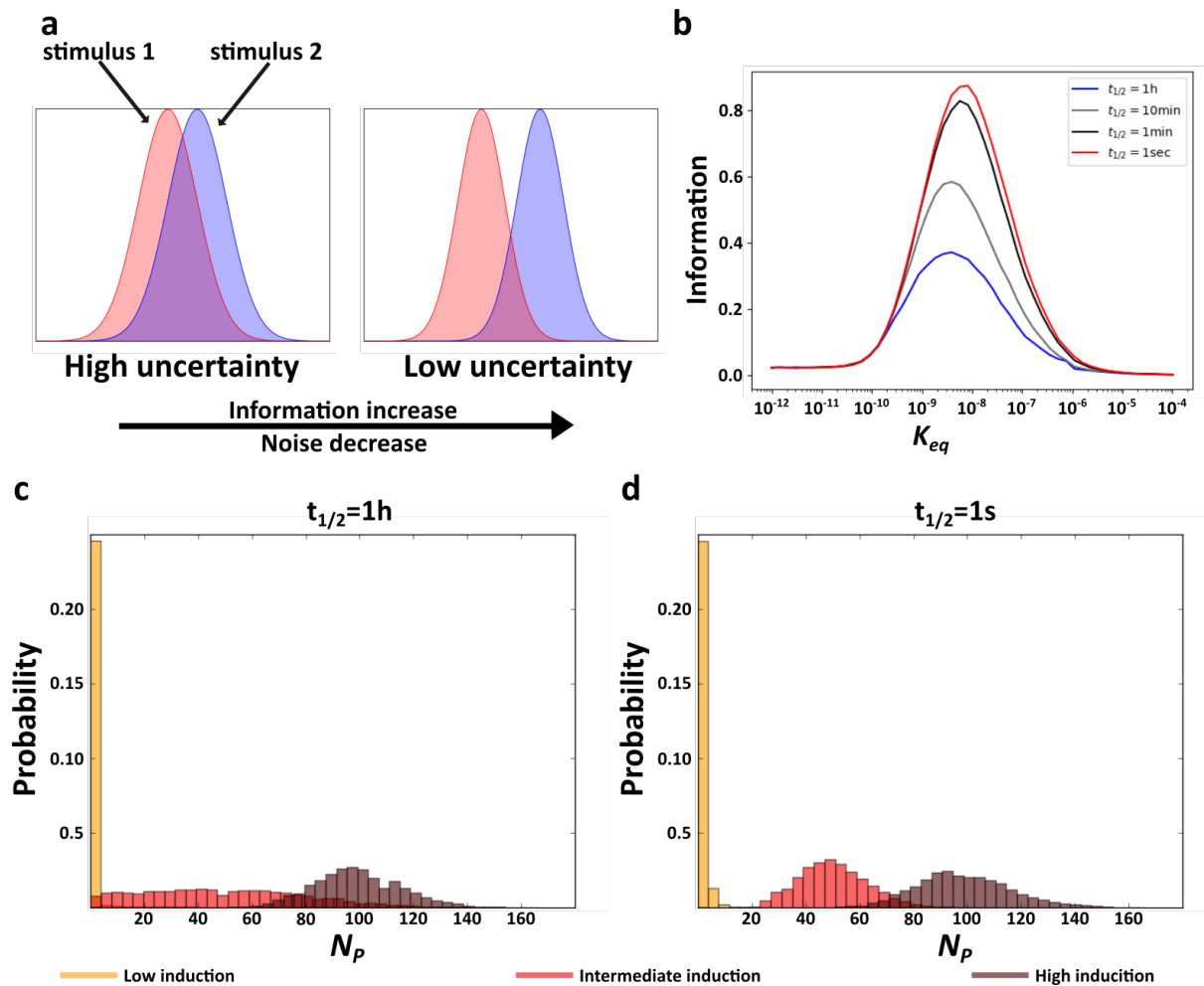
Fig 3. Residence time and information. (a) Schematic explanation of the relationship between noise and acquired information. The panel shows hypothetical response protein distributions produced by two different stimuli at a high (right) and a low (left) level of gene expression noise. (b) Acquired information at different residence times as a function of the affinity (K_{eq}) between TF molecules and *DNA*. (c and d) Distributions of the number of expressed proteins N_P at three different TF concentrations (see color legend) with a long (c), and a short (d) residence time. In c and d, $K_{eq} = 10^{-9}$M.
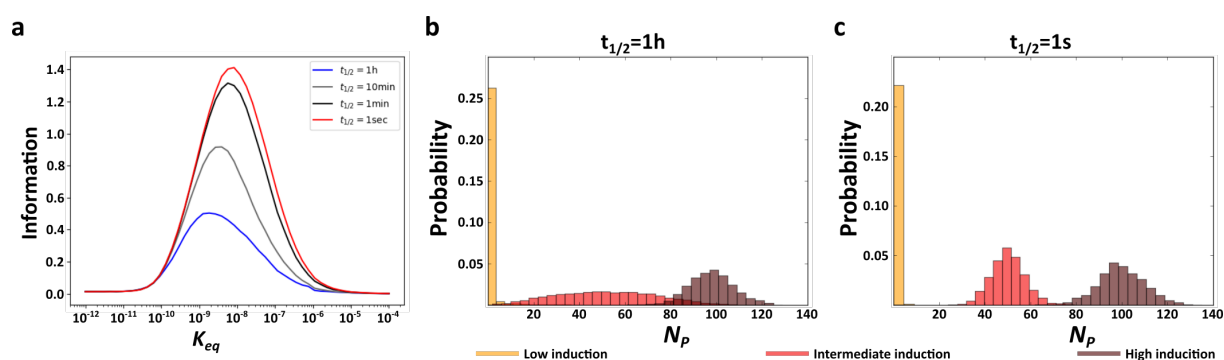
Supplementary figures

16
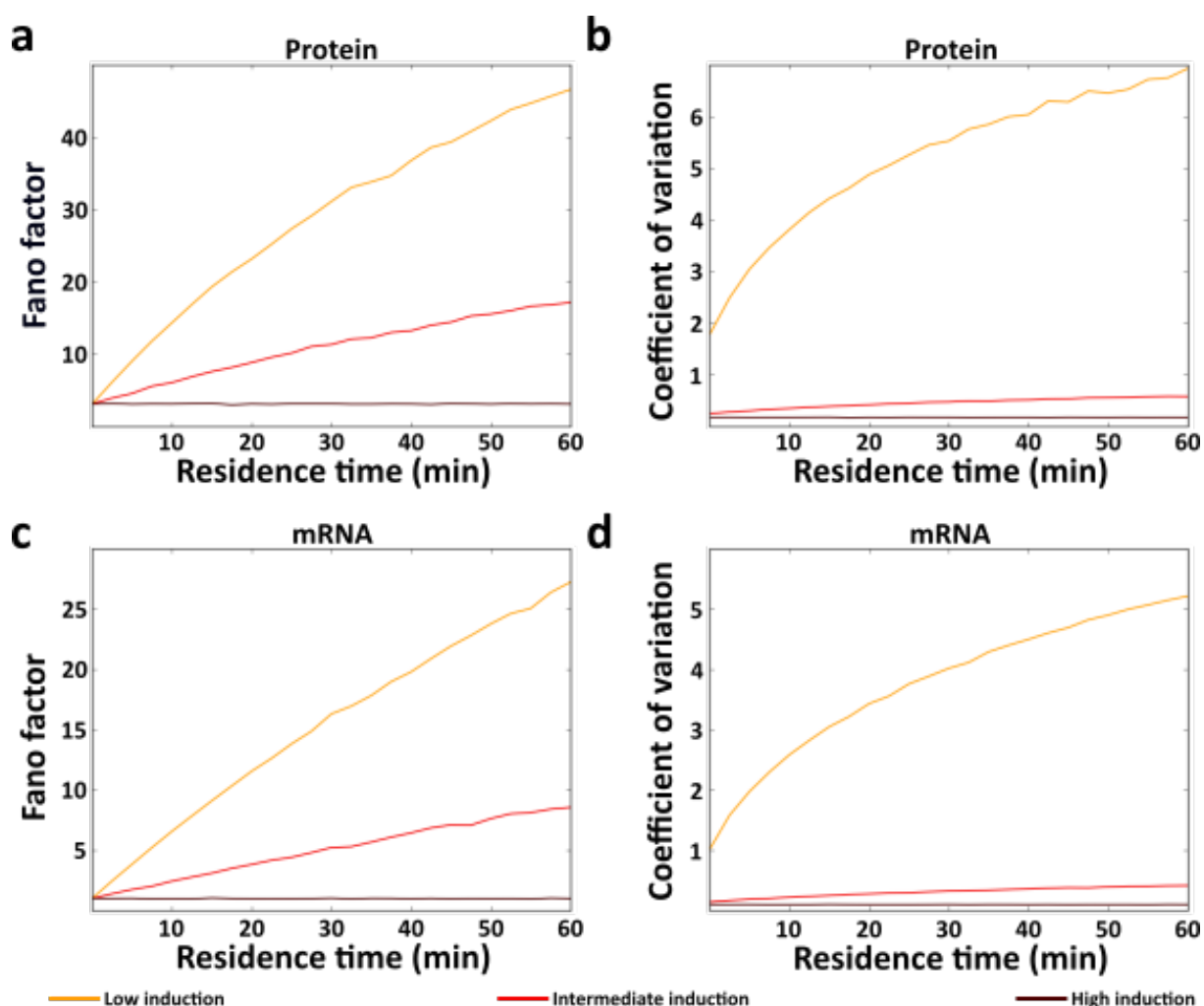
Sup Fig 1. Effect of residence time on Fano factor and coefficient of variation. (a and c) Fano factor and (b and d) coefficient of variation in the number of expressed protein (a and b) and mRNA (c and d) molecules as a function of residence time (*x* axis) at a high (TF=10^{-7}M), intermediate (TF=10^{-9}M) and low (TF=10^{-11}M) levels of induction, as indicated by the color legend below the figure.
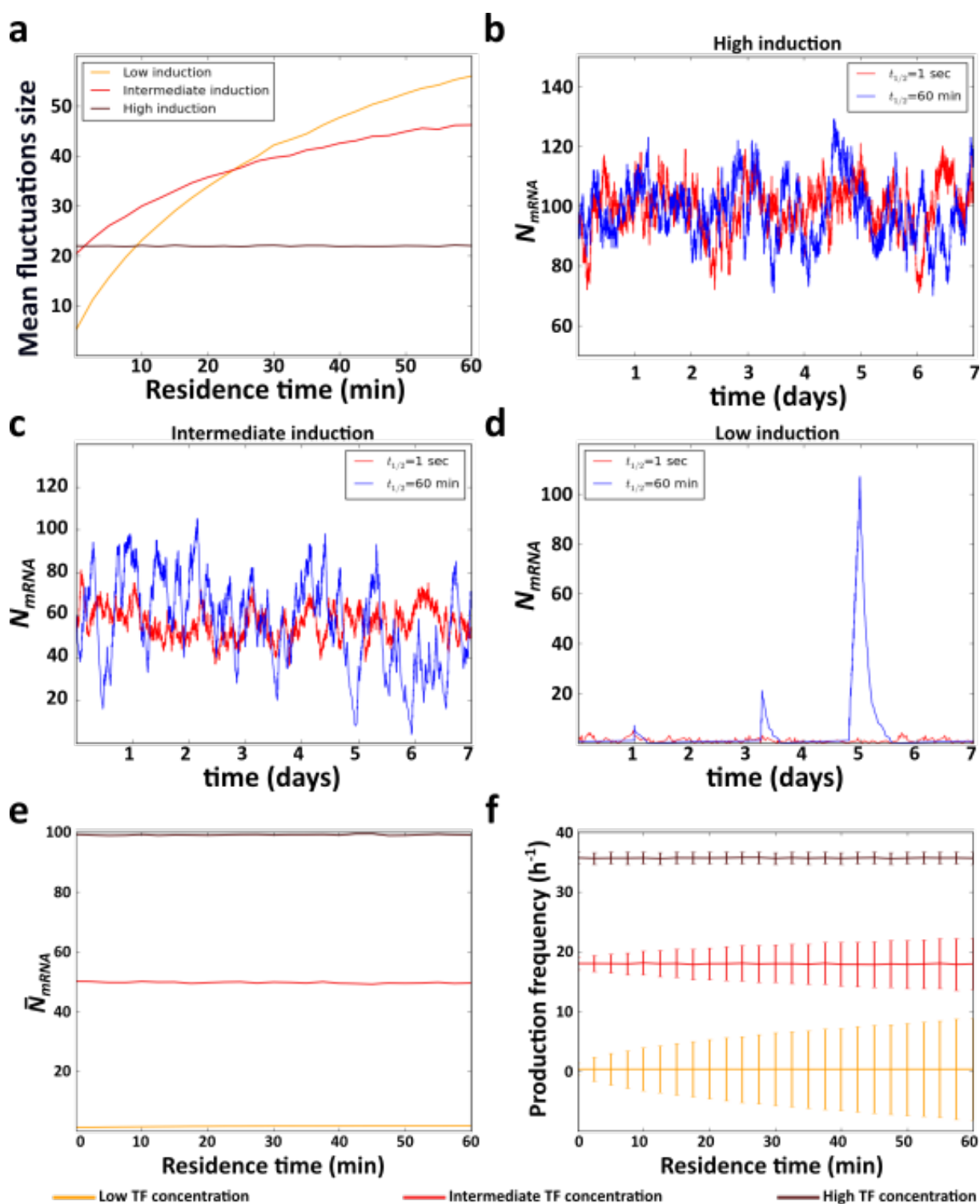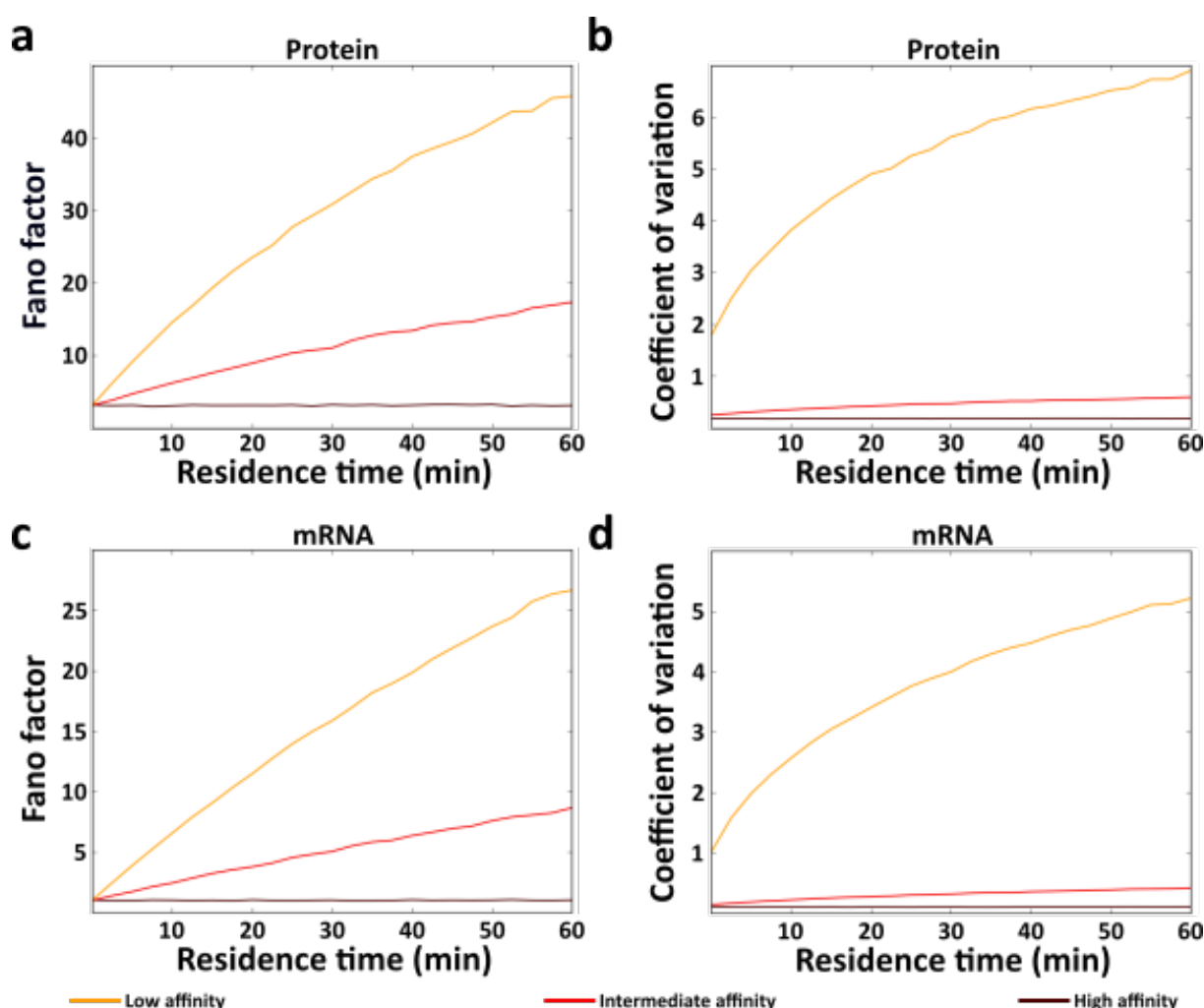
Sup Fig 2. Effect of residence time on the size of the mRNA fluctuations. (a) Mean fluctuation size in the number of mRNA molecules (y axis) at a high (TF=10^{-7}M), intermediate (TF=10^{-9}M) and low (TF=10^{-11}M) level of induction, as a function of residence time (x axis). (b-d) Example time trajectories of the number of expressed mRNA molecules N_{mRNA} obtained from the simulation of the model at three different levels of induction. Red and blue lines show data for short (1s) and long (1h) residence times, respectively. Analyses of (e) mean number of mRNA ($\bar{N}_{mRNA}$) (f) mean and coefficient of variation of the frequency of mRNA production events at high (TF=10^{-7}M), intermediate (TF=10^{-9}M), and low (TF=10^{-11}M) TF concentrations, as indicated in the color legend below the figure.
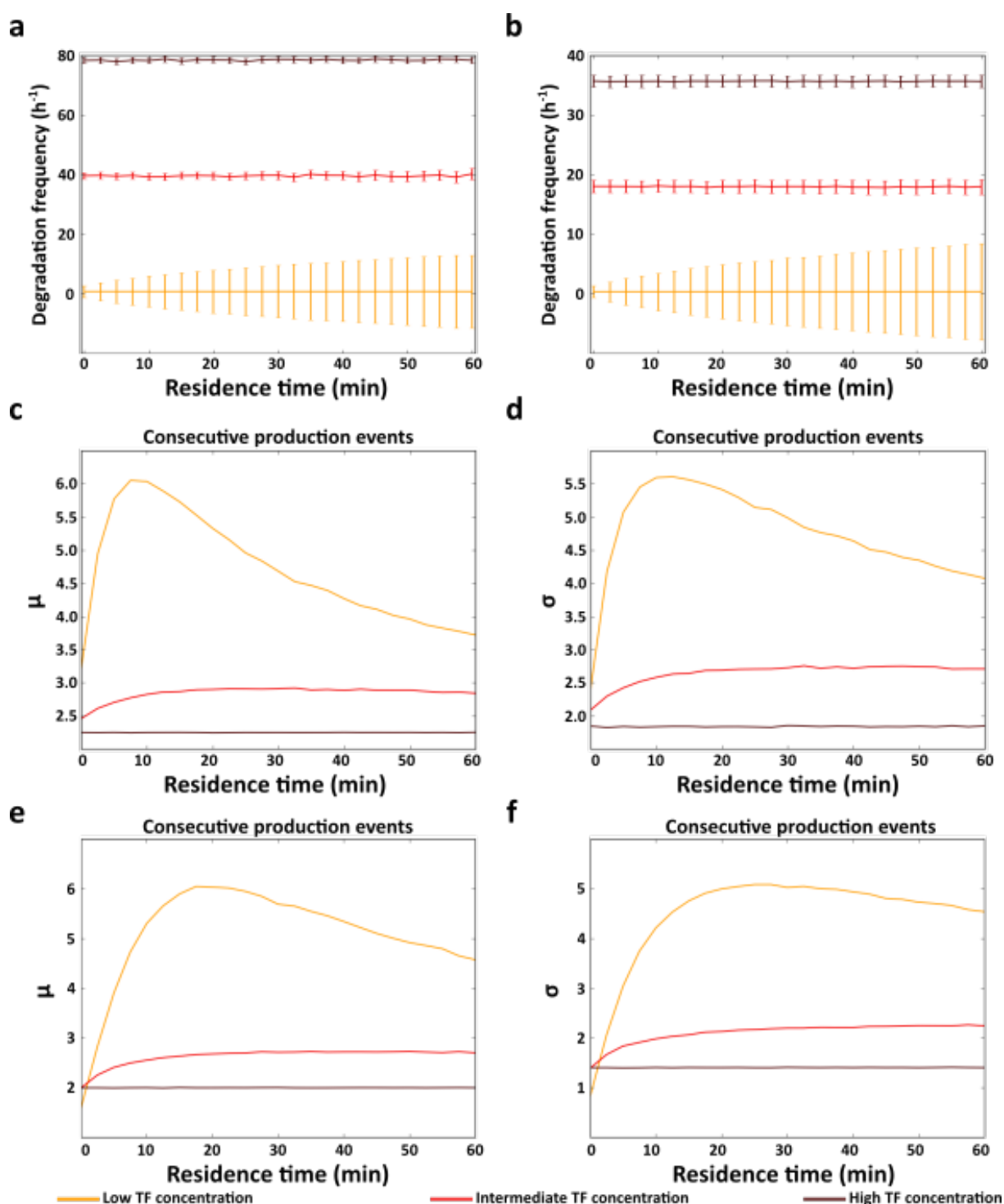
Sup Fig 3. Residence time and both production and degradation events. Mean and standard deviation of (a) protein and (b) mRNA degradation events. (c and e) Mean and (d and f) standard deviation of the number of consecutive (c and d) protein and (e and f) mRNA production events as a function of residence time (*x* axes). All analyses were performed at high (TF=10^{-7}M), intermediate (TF=10^{-9}M) and low (TF=10^{-11}M) affinity values (see color legends)

19

Sup Fig 4. Residence time and acquired information quantified through mRNA expression. (a) Information acquired at different residence times as a function of the affinity (K_{eq}) between TF molecules and *DNA*. (b and c) Distributions of the number of proteins N_P produced at three different TF concentrations with a long (b), and a short (c) residence time. In c and d, $K_{eq} = 10^{-9}$M.


Sup Fig 5. Effect of residence time on Fano factor and coefficient of variation. (a and c) Fano factor and (b and d) coefficient of variation in the number of protein (a and b) and mRNA (c and d) molecules as a function of residence time (*x* axes) at high ($K_{eq}=10^{-11}$M), intermediate ($K_{eq}=10^{-9}$M) and low ($K_{eq}=10^{-7}$M) affinity values, as indicated by the color legend below the figure.

20
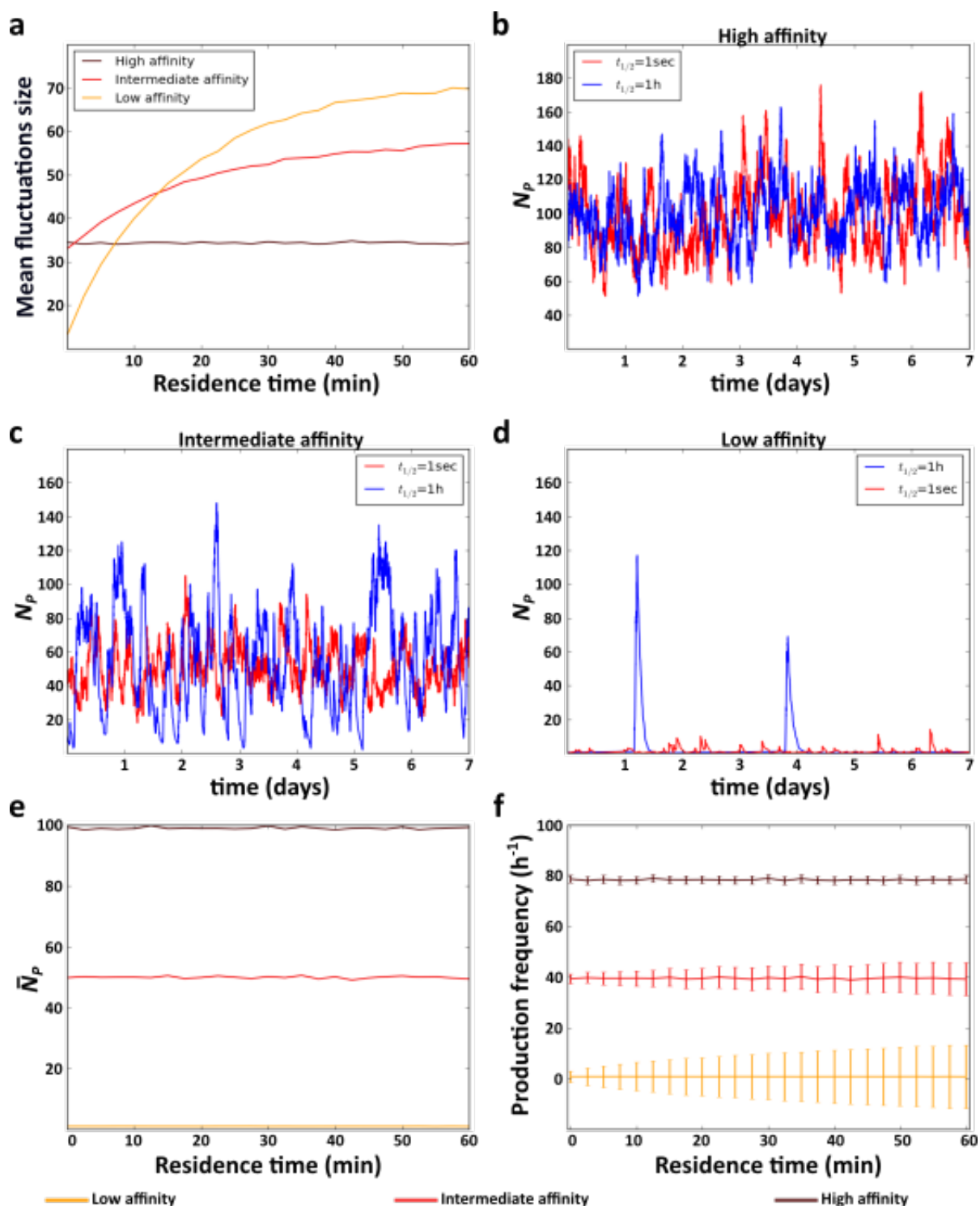
Sup Fig 6. Effect of residence time on the expression dynamic of protein molecules. (a) Mean fluctuation size in the number of protein molecules (*y* axis) at high ($K_{eq}=10^{-11}$M), intermediate ($K_{eq}=10^{-9}$M) and low ($K_{eq}=10^{-7}$M) affinities, as a function of residence time (*x* axis). (b-d) Example time trajectories of the number of protein molecules N_P obtained from the simulation of the model at the three different affinities. Red and blue lines show data for short (1s) and long (1h) residence times, respectively. Analyses of (e) mean number of proteins ($\bar{N}_P$), and (f) mean and coefficient of variation of the frequency of protein production events.
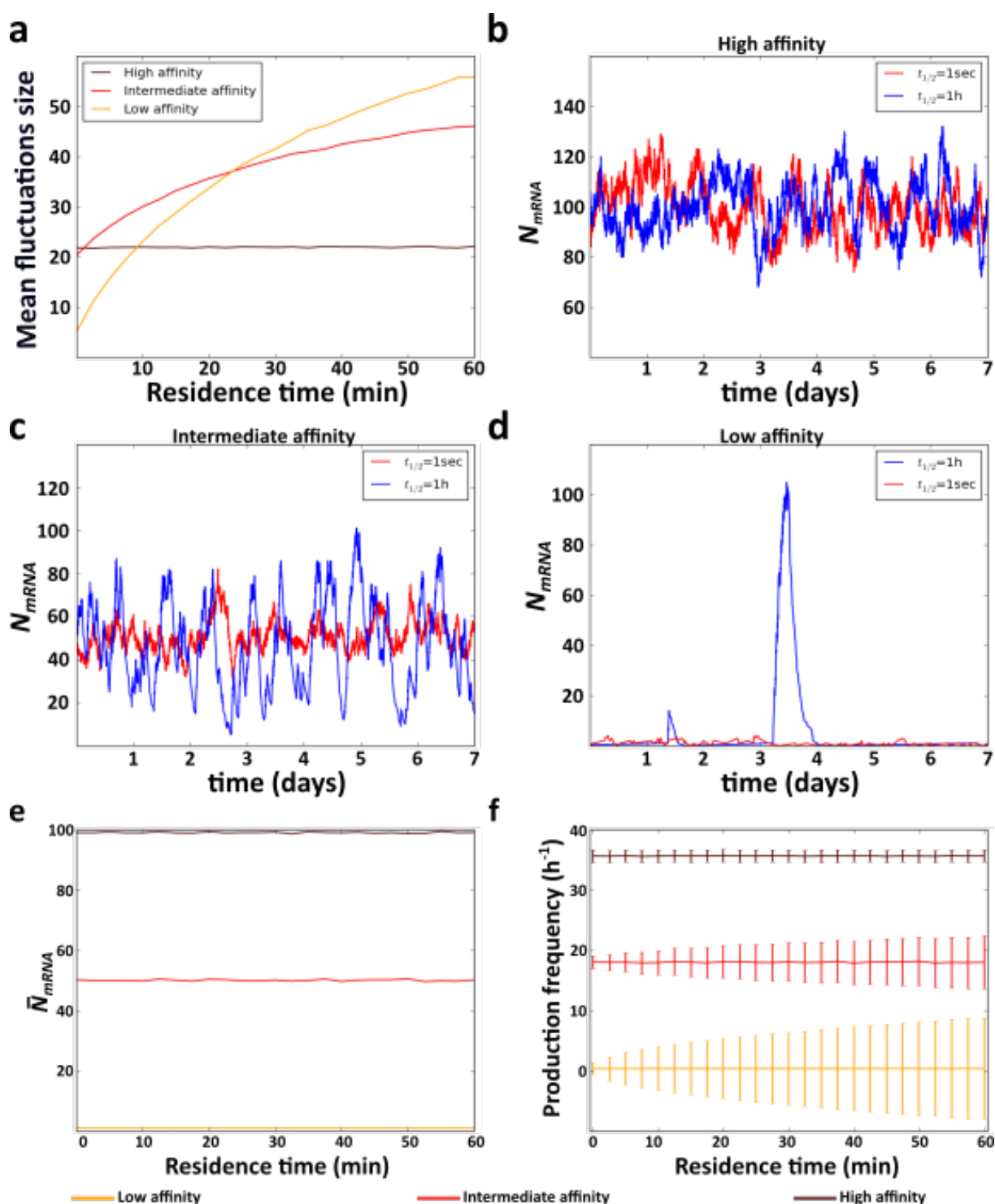
Sup Fig 7. Effect of residence time on time on the expression dynamic of mRNA molecules. (a) Mean fluctuation size in the number of mRNA molecules (*y* axis) at high ($K_{eq}=10^{-11}$M), intermediate ($K_{eq}=10^{-9}$M) and low ($K_{eq}=10^{-7}$M) affinity values, as a function of residence time (*x* axis). (b-d) Examples time trajectories of the number of expressed mRNA molecules N_{mRNA} at three different affinity values. Red and blue lines show data for short (1s) and long (1h) residence times, respectively. Analyses of (e) mean number of expressed mRNA molecules ($\bar{N}_{mRNA}$), and (f) mean and coefficient of variation of the frequency of mRNA production events
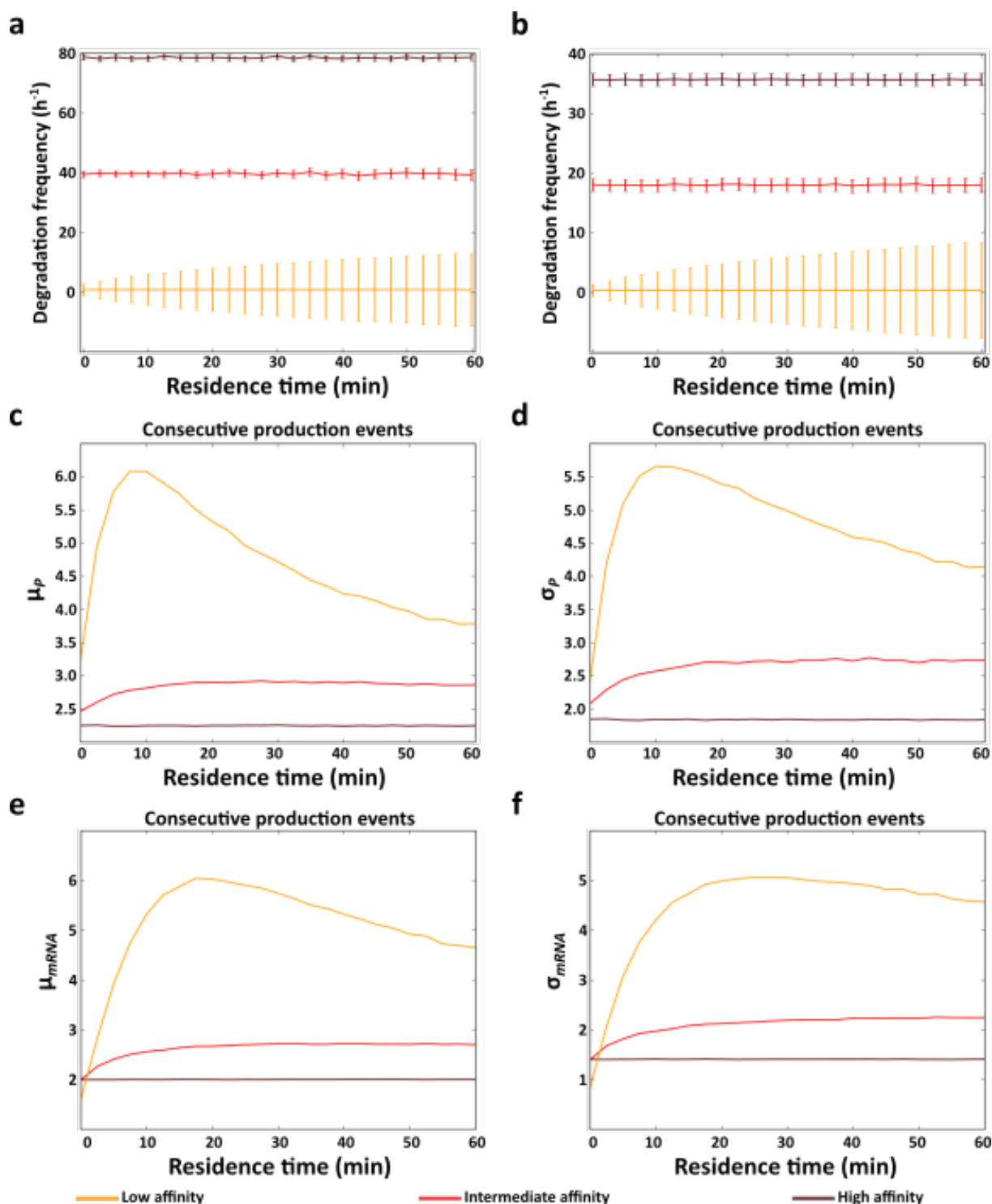
Sup Fig 8. Residence time and both production and degradation events. Mean and standard deviation of (a) protein and (b) mRNA degradation events. (c and e) Mean and (d and f) standard deviation of the number of consecutive (c and d) protein and (e and f) mRNA production events as a function of residence time (x axes). All analyses were performed at high ($K_{eq}=10^{-11}$M), intermediate ($K_{eq}=10^{-9}$M) and low ($K_{eq}=10^{-7}$M) affinity values (see color legends)

Sup Table 1. Parameter values used for the simulations. k_d, k_1, k_2, d_1, d_2 units are s^{-1}. k_a units are M^{-1}s^{-1}. The concentration of the TF is in M.

	k_a	k_d	k_1	k_2	d_1	d_2	TF conc
Fig 2 Sup Fig 1a	$[2.7*10^5, 10^8]$	$[2.7*10^{-4},1]$	0.01	0.011	0.005	0.00022	$[10^{-11},10^{-7}]$

and b Sup Fig 3a, c and d							
Fig 3	[2.7, 1.0^{12}]	[$2.7*10^{-4}$,1]	0.01	0.011	0.005	0.00022	[10^{-11},10^{-7}]
Sup Fig 1c and d Sup Fig 2 Sup Fig 3b, e and f	[$2.7*10^5$, 10^8]	[$2.7*10^{-4}$,1]	0.01	NA	0.0001	NA	[10^{-11},10^{-7}]
Sup. Fig 4	[2.7, 1.0^{12}]	[$2.7*10^{-4}$,1]	0.01	NA	0.0001	NA	[10^{-11},10^{-7}]
Sup Fig 5a and b Sup Fig 6 Sup Fig 8a, c and d	[$2.7*10^3$, 10^{10}]	[$2.7*10^{-4}$,1]	0.01	0.011	0.005	0.00022	10^{-9}
Sup Fig 5c and d Sup Fig 7 Sup Fig 8b, e and f	[$2.7*10^3$, 10^{10}]	[$2.7*10^{-4}$,1]	0.01	NA	0.0001	NA	10^{-9}

Supplementary information

1. Affinity and induction level

As explained in the main text, the affinity of a TF to its DNA binding site and the TF concentration determine the induction level of a gene (Fig 1d). Hence, we hypothesized that the effect of affinity on gene expression could be reproduced by a change in the TF concentration. To find out, we repeated all our analyses, but instead of varying the affinity, we set it to a constant value of 10^{-9}M. We then varied the TF concentration within the interval [10^{-7}M,10^{-11}M]. Analogous to our analysis in the main text, this interval includes TF concentrations two orders of magnitude above and below the affinity. Hence, at the highest TF concentration, the level of induction is high and the regulated gene is almost always active. Conversely, at the lowest TF concentration, the level of induction is low and the regulated gene is almost always inactive. We were able to reproduce all observations we had made by varying affinity through the variation of TF concentrations. Most importantly, as the level of induction decreases, shorter residence times reduce noise (compare Fig 2, Sup Fig. 1 and Sup Fig 2 with Sup Fig. 5-7), producing a more homogenous and regular dynamic of gene expression (compare Sup Fig. 3 with Sup Fig. 8). Hence, our observations show that affinity affects gene expression dynamics mainly through its effect on gene induction.

2. Residence time and strength of noise

To study the effect of residence time and affinity on the level of gene expression noise, we also quantified the Fano factor, and the coefficient of variation of gene expression. The Fano factor is equal to the variance in the number of expressed molecules divided by its mean ($\sigma^2(N_P)/\overline{N}_P$), while the coefficient of variation is equal to the standard deviation

divided by the mean ($\sigma(N_P)/\overline{N}_P$). Both measures provide information about the dispersion of the distribution of expressed molecules relative to its mean.

The Fano factor and the coefficient of variation display the same behavior as the standard deviation in the number of expressed molecules (i.e., the size of the fluctuations). When a gene is highly induced at the highest affinity value, residence time does not affect any of these quantities (Sup Fig. 1). As explained in the main text, the reason is that the regulated gene behaves like a constitutive gene in this case, where the fluctuations of protein concentrations around their mean do not depend on a TF's residence time (Fig. 2a and d; Raj and van Oudenaarden 2008).

At the opposite extreme, as induction approaches zero, longer residence times increase the Fano factor and the coefficient of variation. The reason is that longer residence times produce larger fluctuations in the number of expressed molecules (Fig. 2a). Not surprisingly then, the variance in protein number is higher for longer residence times, which increases the strength of noise (Sup Fig. 1), because the mean protein and mRNA expression is not affected by residence time (Fig. 2e; Sup Fig. 2e). In sum, similar to our results in the main text, as the level of induction decreases, the amount of noise decreases with residence time.

3. Protein and mRNA production and degradation events

Because shorter residence times (at submaximal induction) reduce variation in the frequency of protein production and degradation events, we hypothesized that production and degradation events should alternate more regularly at short residence times, such that the number of consecutive production events (i.e., the number of protein production events without any intervening degradation event should decrease). To validate this hypothesis, we quantified the number of consecutive production events at different levels of induction and residence times. When induction is high, residence time does not affect the number of consecutive production events (Sup. Fig. 3c and e), because gene expression resembles that of a constitutive gene at all residence times (Fig. 2a and b). However, as induction decreases, shorter residence times decrease both the mean and standard deviation number of consecutive production events (Sup Fig. 3c-f), showing that production and degradation events alternate more frequently.

One exception to this pattern occurs at the lowest level of induction. Here, the mean and standard deviation of the number of consecutive production events reaches a maximum at intermediate residence time and decreases as the residence time increases (Sup Fig. 3c-f, yellow). We believe that the reason is that most production events tend to occur in a few but long periods of active gene expression when the induction level is low and the residence time is long. During these long periods of active gene expression, protein expression behaves as for a highly induced gene, which has a smaller mean and standard deviation of the number of consecutive production events (Sup. Fig. 3c-f). Notice, however, that this dynamic will still produce high levels of noise, because mRNA and protein molecules will fluctuate between very low and very high levels (Fig 2a; Sup Fig. 2a). In conclusion, the more frequent alternation of production and degradation events observed at shorter residence times produce a more regular and less noisy dynamic of expression.