Master Thesis

Quantifying in Vitro Transcription

T.P.Visser (5644550)

Under supervision of:
W.K.Kegel
G.Folkers

Van 't Hoff Laboratory for Physical and Colloid Chemistry
Debeye Institute for Nanomaterials Science
Utrecht University
September, 2022

Contents

1 Abstract

Transcription of DNA to mRNA has been studied extensively, but quantifying the rate of transcription often remains a challenge. We explore a new method to study in vitro transcription by using a specific DNA sequence, which translates to mRNA that is able to bind to a fluorescent dye. With this method, the fluorescence of mRNA can be measured directly in time and the transcription rate can easily be determined, for multiple processes at once. We combine data obtained by our experiments with a thermodynamic model. The model predicts the relative transcription rate by calculating the occupation of transcription sites at determined concentrations of polymerase and DNA. We find that the binding energies we use for theoretical model fall within the range of values in literature and are consistent with independent binding assays. We conclude that more complex systems with additional transcription factors, such as lac suppressor, can also be studied. Being able to quantitatively study in vitro transcription opens up a versatile test ground for models of gene regulation.

2 Introduction

Gene regulation is of major interest in molecular biology as it the first step in a complex multi-step process which eventually leads to gene expression. Understanding gene regulation means being able to better understand how certain genes come to expression and what exactly affects gene expression. In biological cells, genes code for proteins. The production of proteins is regulated by a complex regulatory network involving regulatory proteins (transcription factors). Each gene competes for transcription factors with a complex array of binding sites on the genome. Genes often exist in multiple identical copies, for example in the process of chromosomal replication during the cell cycle. Predicting the quantitative effect of this competition on the regulation of gene expression is of great interest. There are many different ways of approaching quantitative gene regulation [1], [10], [22], [25], [4] [30]. In this project we make use of a model introduced by Weinert et al[30] and expanded on by Landman et al [15]. In this model the grand canonical ensemble is used to construct a statistical mechanical probability that a transcription factor is bound to its specific site. To measure transcription we make use of a sequence called Broccoli [8]. Broccoli allows us to measure transcription quantitatively in real time. The model is then adapted to the experimental system and compared to the experimental measurements.

3 Theory

The first step in gene expression is transcription. Transcription is the process of reading the DNA sequence and formation of mRNA by action of RNA-polymerase (RNAP). This process is initiated by the binding of RNAP to the promoter sequence of the gene of interest on the DNA. The binding probability of RNAP to the promoter depends on the number of RNAP available to the

gene, the binding energy, the number of gene copies and the number of transcription factors that may either suppress or promote binding of RNAP to the promoter. [15] The first is a 'repressor' which adsorbs to the promoter region and prevents adsorption of RNAP by excluded volume. The second is an 'activator' that may bind to RNAP and an operator region close to the promoter, effectively increasing the affinity of RNAP to the promoter.

The model used in this work is based on several key assumptions. The most critical assumption is quasi-equilibrium, this is the assumption that the probability of finding the polymerase on the promoter is given by its equilibrium value. This is only possible if there is sufficient time separation between the various processes that occur during transcription; RNAP binding, open complex formation, transcript formation and transcription itself. [15]

The number of specific sites or promoter sites is referred to as N_s. The subscript "s" stands for specific, subscript "ns" for nonspecific sites and subscript "c" for competitor sites. When RNAP binds to a promoter site it is activated and starts the transcription process. For T7 the promoter site is a sequence of around twenty base-pairs. A promoter site can also have a binding site for a repressor, such as LacI, depending on the gene structure. When LacI binds to a promoter, RNAP is prevented from activating, thus inhibiting transcription. There exist many other regulatory structures, but these will not be discussed in this project.

Non-specific sites are all the binding sites on the DNA sequence that do not specifically bind to a transcription factor. Transcription factors can still bind to non-specific DNA and often does, which has an effect on the transcription rate. All binding energies in the model are measured and calculated with respect to the free energy of the respective transcription factor dissolved in water.
Fold Change is used to quantify transcription

$$FC = \frac{< P_s >}{< Ps_{ref} >} \tag{1}$$

Where $< Ps >$ is the average number of bound RNAP to promoter sites and $< Ps_{ref} >$ is the number of bound RNAP in a reference state.

The promoter sites can have any number of states, each of which has a specific grand canonical weight based on the free energy of that state. In this case we regard a promoter that can be bound by RNAP or unbound. The grand canonical partition function is then given by:

$$\Xi_s = (1 + \lambda_p e^{-\beta \epsilon_p})^{N_s} \tag{2}$$

Where ϵ is the binding energy in k_BT and $\beta = \frac{1}{k_B T}$. λ_p is the fugacity of the polymerase given by: $\lambda_p = e^{\beta \mu_P}$ with μ_p the chemical potential of the polymerase.

The fugacity is one of the most important quantities in this model. The fugacity is the effective concentration of RNAP. If the fugacity and binding energies of all sites is known, the occupation of any binding site in the system

can be calculated via:

$$< P_s >= \lambda_p \frac{\partial ln \Xi_s}{\partial \lambda_p} = N_s \frac{\lambda_p e^{-\beta \epsilon_p}}{1 + \lambda_p e^{-\beta \epsilon_p}} = N_s \theta \tag{3}$$

θ is the occupation of the binding site. The occupation of the specific sites the key quantity that provides a connection to the experimental transcription rate.

$$\theta = \frac{\lambda_p e^{-\beta \epsilon_p}}{1 + \lambda_p e^{-\beta \epsilon_p}} \tag{4}$$

The average number of bound RNAP can be constructed for any type of binding site for which the grand canonical partition function can be constructed. As the total number of RNAP molecules must be conserved, the fugacity can be calculated from the mass balance. The total copy number P of RNAP must be equal to the number of bound RNAP to all the possible binding sites in the system:

$$P =< P_s > + < P_{ns} > + < P_{aq} > \tag{5}$$

Where $< P_s >$ is the number of the RNAP bound to specific sites, $< P_{ns} >$ to non-specific sites and $< P_{aq} >$ the number dissolved in water. For each of these sites the grand canonical partition function is set up and inserted in equation 3. Applying this to the mass balance above for a system with just RNAP yields:

$$P = N_s \theta_s + N_{ns} \theta_{ns} + < P_{aq} > \tag{6}$$

Where $\theta_i = \frac{\lambda_p e^{-\beta \epsilon_{pi}}}{1 + \lambda_p e^{-\beta \epsilon_{pi}}}$ It is practical to divide this equation by the volume of the system V and Avogadro's number N_{Av}, converting the copy numbers to molar concentrations.

$$[P] = [N_s] \theta_s + [N_{ns}] \theta_{ns} + \frac{< P_{aq} >}{V N_{Av}} \tag{7}$$

This last term $\frac{<P_{aq}>}{V N_{Av}}$ is now the effective concentration of RNAP in water and equal to the fugacity λ_p. Note that the dimension of the fugacity is now in molar concentration rather than mole fraction like in the work of Landman et al [15]. This also has consequences for the binding energy, as will be discussed later. Thus we can rewrite to:

$$[P] = [N_s] \frac{\lambda_p e^{-\beta \epsilon_{ps}}}{1 + \lambda_p e^{-\beta \epsilon_{ps}}} + [N_{ns}] \frac{\lambda_{pns} e^{-\beta \epsilon_{pns}}}{1 + \lambda_p e^{-\beta \epsilon_{pns}}} + \lambda_p \tag{8}$$

The only unknown in this equation is the fugacity of the RNAP λ_p. By solving for λ_p the occupation of all the binding sites in the system can be evaluated.

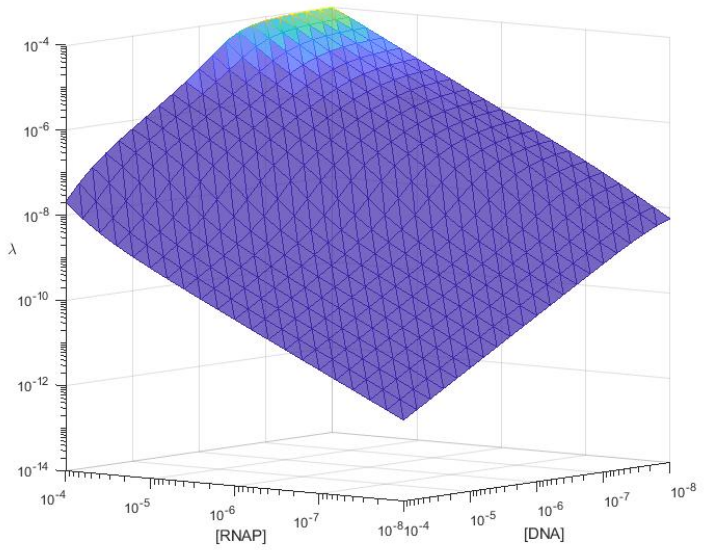

Figure 1: Log log plot of the fugacity λ as a function of the concentration in mole/L of RNAP and DNA

In figure 1 a 3D plot of the fugacity shows that as RNAP is added to the system, there are two shifts in the steepness of the curve, one at low concentration where the specific sites are saturated and one at a high concentration where the non-specific sites are saturated. In figure 2 the effect of the RNAP concentration is plotted in 2D to give a clearer picture.

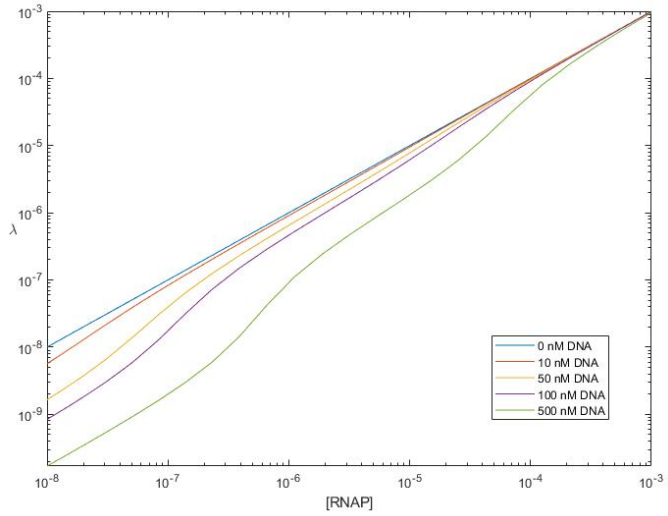

Figure 2: 2D log log plot of the fugacity λ as a function of the concentration of RNAP in mole/L at several different concentrations of DNA

These two shifts are made more visible in figure 2. These shifts corresponds to the saturation of the specific and nonspecific binding sites as a result of the titration effect. The concentration of RNAP at which these transitions take place in related to the binding energies corresponding to the site in question. Eventually, when enough RNAP is added to the system, the fugacity starts to approach the total concentration of RNAP.

In molecular biology, the binding energy if often expressed as the dissociation constant or K_d. The dissociation constant or K_d is the ratio between the association rate K_{on} and the dissociation rate K_{off}.

$$K_d = \frac{k_{off}}{k_{on}}$$

The dissociation constant can therefore be determined by kinetic measurements of the k_{on} and k_{off}. The K_d can also be expressed as a concentration as it is the concentration at which half of a certain molecule is associated with a substrate at equilibrium.

$$K_d = \frac{[A][B]}{[AB]}$$

Where A, B and AB are the equilibrium concentrations of A, B and the complex AB. From this it follows that when $[A] = K_d$, $[B] = [AB]$ or half of the substrate B is part of a complex AB. This is the equilibrium dissociation constant, which is what will be used to calculate the binding energy of the sites in the system.

The transitions visible in figure 1 and 2 occur when the fugacity is equal to the K_d of the corresponding binding sites in the system. The K_d is related to

the binding energy by $\epsilon_p = k_B T ln(\frac{K_d}{\phi})$ where ε_p is the difference between the free energy of an RNAP molecule bound to its specific site compared to free in solution. ϕ is the standard reference concentration of 1 mole/L, to convert the Kd expressed in concentration to a unit-less constant. The standard reference concentration is thus used as the reference state to calculate the equilibrium binding energies of the different sites.

In the work of Landman et al, it is generally assumed that all the polymerase is bound to a site on the DNA. In their work the fraction of DNA dissolved in the water does not need to be taken into account. The natural reference state is polymerase bound to the binding sites with the lowest binding energy, the non-specific sites. The binding energy of any site in the system is then the difference between the binding energy to the non-specific sites and the binding energy to the site of interest. In our case there is a need to take the polymerase dissolved in the water phase into account. This is due to an excess of T7 polymerase in our experiments, which means there is a significant amount of polymerase still dissolved in the water. The T7 polymerase used in our experiments also dissolves quite well in water, which means that even at a relatively low ratio of polymerase to DNA, there is still a portion dissolved in the water. For this reason our reference state is the polymerase dissolved in water at the standard concentration of 1 mole/L.

Competitor sites are sites that have a strong binding affinity towards RNAP and can therefore compete with specific sites. In most cases competitive sites have a stronger binding affinity than specific sites. As a consequence competitive sites will be favored over specific sites, effectively inhibiting transcription by binding RNAP. We can add an additional term to the mass balance in equation 5 to account for competitor sites:

$$P =< P_s > + < P_{ns} > + < P_c > + < P_{aq} >$$

With $< P_c >$ the number of RNAP bound to competitive sites.
Filling in equation 3 in the equation above yields.

$$[P] = [N_s]\frac{\lambda_p e^{-\beta\epsilon_{ps}}}{1+\lambda_p e^{-\beta\epsilon_{ps}}} + [N_{ns}]\frac{\lambda_{pns} e^{-\beta\epsilon_{pns}}}{1+\lambda_p e^{-\beta\epsilon_{pns}}} + [N_c]\frac{\lambda_p e^{-\beta\epsilon_{pc}}}{1+\lambda_p e^{-\beta\epsilon_{pc}}} + \lambda_p \quad (9)$$

This results in a quartic equation for λ_p, which can still be solved easily with a given number of competitor sites and binding energy ϵ_{p_c}.

4 Method

In vitro transcription is a useful tool to study gene regulation [19]. The conventional method of measuring in vitro transcription is RNA radio labelling followed by denaturing PAGE [17]. This method is unfortunately very labor intensive and cannot be done in vivo. In this work we make use of a method to measure in vitro transcription, which should be viable in vivo as well. For this purpose we use Broccoli [8]. Broccoli is a DNA sequence that forms Broccoli mRNA when transcribed. This mRNA sequence folds itself into a non-fluorescent aptamer. This aptamer binds to a dye, DFHBI-1T to form a highly fluorescent complex. Broccoli has been succesfully used to study in vivo gene expression qualitatively [23], [2], [16]. In vitro, Broccoli can be used as a direct fluorescent readout for transcription [14]. A similar sequence, known as Spinach, can be also used to measure transcription in vitro. [7]

The double stranded DNA fragments used in these measurements contain a promoter sequence, which serves as the specific site for T7 RNA polymerase, the broccoli sequence and when stated an O1 [25] site, which serves as the specific site for a lacl repressor. The fragments without O1 site have a length of 140 base pairs and with the O1 sequence 170 base pairs.

In vitro transcription was performed using T7 RNA polymerase based on ThermoFischer Scientific^{TM} protocol [27] with some adjustments. Reactions were conducted in 25 µl instead of 50 µl and the DNA concentration ranges from 0.25 ng to 44 ng instead of 500 ng. No RNase inhibitor was used. 0.5 µl 200 mM DFHBI-1T was added to the reaction. 100 ng/µl BSA was added when indicated. Reactions were carried out in a Greiner 384 wells plate. Fluorescence was measured with a FLUOstar OPTIMA plate reader or a SpectraMax iD3 Multi-Mode Microplate Reader depending on the experiment. Measurements lasted one to two hours and fluorescence was measured every 60 seconds to produce a fluorescence over time data. The reaction rate is then determined by linear fit of the first 30 minutes of the reaction, an example of such a fit can be found in appendix A.

5 Results and discussion

5.1 Fitting the data

We compare the model with experiment by assuming the occupation of the specific sites is directly proportional to the reaction rate. To be able to compare the transcription rate with the model, a linear fit is made between the experimental transcription rate versus the theoretical occupation ratio, see figure 3. For each concentration of RNAP, the occupation is calculated and compared to the experimental transcription rate. This is done separately for each experiment as the measured fluorescence has an arbitrary intensity that can vary from experiment to experiment. For experiments with several concentrations of DNA, the series with the highest concentration of DNA is taken as reference.

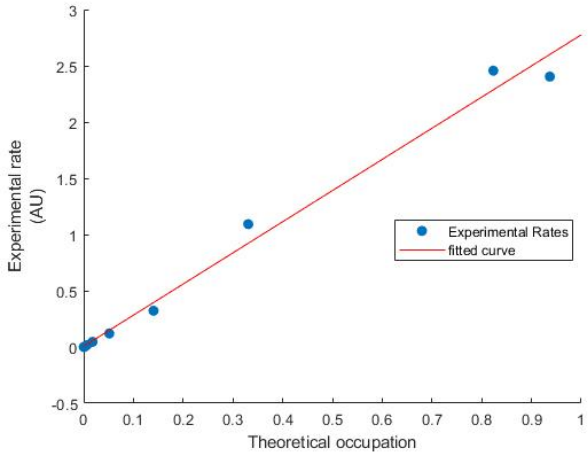

Figure 3: Reaction rate in AU on the y axis versus the theoretical occupation θ as calculated by the model on the x axis.

To fit the lower concentrations of DNA the fit parameter is normalized with the relative amount of DNA in the system. The data with high concentrations of DNA has the highest activity and least noise and therefore performs most consistently. If there is a point in the series that significantly deviates from the set, that point is taken out as an outlier to improve the fit. This is worth noting as in some experiments, the highest concentrations of RNAP are significantly less active than lower concentrations, see appendix B. The reason for this decrease in activity at the highest concentrations of RNAP is unclear, but could be caused by aggregation of the enzyme.

5.2 Finding the binding energy

To be able to model a system and compare with experimental reaction rates, it is necessary to know the equilibrium binding energy of the specific sites. The equilibrium binding energy can be calculated directly from the equilibrium dissociation constant.

The formation of a transcription capable complex is a complicated process with several steps which all have their own dissociation constants. First, the RNAP must find and recognize the promoter DNA sequence through specific protein-DNA interactions. This specific, promoter-bound complex is known as the "closed" complex. [6] After promoter recognition, the DNA helix undergoes RNAP-induced bending and melting in the promoter region to form the "open" complex. With a single initiating ribonucleotide (NTP), abortive RNA synthesis starts with the release of short RNA chains with a length of 7–12 NTP. [18] After formation of the stable initial transcription complex, consisting of the short RNA transcript, DNA template, and RNAP, promoter escape and transition to a stable elongation complex occurs. [26] The combination of these steps gives rise to the equilibrium dissociation constant, which determines the number of active RNAP at a given concentration in a steady state system.

The equilibrium dissociation constant of T7 RNAP to the T7 promoter was investigated with fluorescence anisotropy (FA) [20] and flow-induced dispersion analysis (FIDA) [24]. Several measurements were performed to find the dissociation constant of T7 RNAP towards the specific and non-specific sites. The KD of RNAP to its specific site was measured to range from 400 to 600 nM with FA and 84-430 nM with FIDA. This is a broad range of results for the dissociation constant, but still within the same order of magnitude. The difference in binding energy between lowest and highest measured value, 84-600 nM, is around 2 k_BT which is thermodynamically significant.. For the FA measurements, the difference in k_BT between 400 and 600 nM is 0.406 k_BT.

Figure 4 shows the FA measurements and analysis of the T7-O1 DNA sequence that was used for the majority of the experiments. The FA results were fitted with a simple absorption isotherm to find the dissociation constant KD:

$$A * \frac{x}{[Kd] + x} + b \tag{10}$$

Where b is the initial background value, A the maximum value and x the concentration of RNAP.

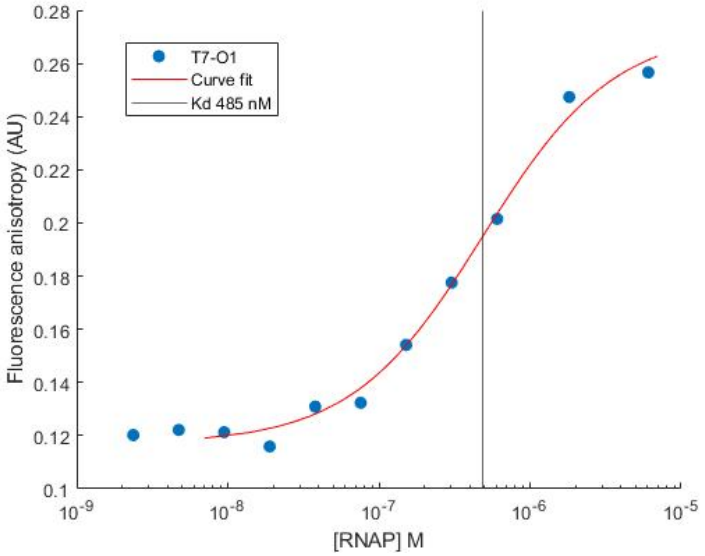

Figure 4: Lin Log plot of the fluorescence anisotropy of T7-O1 DNA as a function of the molar concentration of RNAP. Dissociation constant is indicated with the vertical line.

From this fit, the Kd was found to be 485 nM. From this result, the specific binding energy was calculated to be 14.5 k_BT. This is the value is used in the model to calculate the theoretical occupation of the promoter sites according to equation4. The impact of varying the specific binding energy on the model was investigated by calculating the correlation coefficient of several binding energies between 12.5 and 16.5 k_BT, the results of which can be found in table 1. Corresponding plots of the fits can be found in Appendix C. From this table it is clear that the binding energy of 14.5 k_BT results in the best fit. This means that the dissociation constant measured using FA is consistent with the binding energy used to calculate the occupation of specific sites for the rate experiments.

Binding energy in k_BT	R	R^2
12.5	0.895	0.801
13.5	0.947	0.897
14.5	0.973	0.947
15.5	0.964	0.929
16.5	0.938	0.880

Table 1: Specific binding energies with their corresponding correlation coefficient R and the coefficient of determination R^2

In literature, the dissociation constant of the initiation step in the binding process varies from K_d of 5 to 15nM [13] [29] [21], these dissociation constants are constructed from kinetic rates. The consensus is that the rate limiting step in T7 transcription is not the binding of the promoter or the open complex formation, but rather a transition step between initiation and elongation. A single molecule study identified a rate limiting step between initiation and elongation of the T7 RNAP-DNA complex [26]. This study found that the rate at which the RNAP escapes the promoter and continues to elongation is much slower than the dissociation rate between promoter and RNAP. Furthermore, for complexes that enter the elongation stage, there is a significant lag period before promoter escape. In studies that measured the equilibrium K_d of complex formation between T7 RNAP and DNA, the value ranges from 220 nM [28] to 0.5 µM [3] [11]. These measurements of the equilibrium dissociation constant are of the same order of magnitude as our measurements with FA and FIDA.

Fluorescence anisotropy measures the change in rotation by measuring the change in polarization, as the RNAP binds to the DNA the rotation of the linear DNA molecules is slowed down. [20] The dissociation constant can be determined by labeling the DNA with fluorescent reporters and measuring the change in anisotropy with increasing RNAP concentration. The K_d can then be found by fitting the anisotropy, which should be where half of the RNAP is associated with the DNA. This assumes that the anisotropy changes linearly with the amount of bound RNAP.

FIDA measures the change in diffusion as a result of complex formation. [12] Large molecules diffuse slowly, and small molecules diffuse comparatively faster. When a molecule binds to a ligand its apparent diffusivity becomes that of the formed complex. The apparent diffusivity is dependent on the fractions of free and complexed DNA which in turn is given by the DNA-RNAP equilibrium binding constant. The apparent dispersion of an analyte in the flow system is characterized by the apparent diffusivity of the analyte. FIDA measurements result in a fluorescence intensity over time graph. A concentration gradient forms in the flow, due to differences in diffusivity, this gradient results in a peak, which is fitted to a Gaussian. The width, appearance time and variance of the Gaussian can be used to find the apparent diffusivity of the complex and from there the hydrodynamic radius can be calculated. The fraction of (un)bound DNA can be found from the change in hydrodynamic radius due to complex formation as RNAP binds to the DNA. This fraction can then be used to calculate the dissociation constant. [24]

Both of these methods measure complex formation indirectly. In principle both techniques measure the equilibrium binding constant and should result in the same value, in this case the difference could be due to the poor data quality from FIDA measurements. With fluorescence anisotropy the dissociation constant is found by fitting the half maximum anisotropy. This is done by taking the minimum anisotropy and the maximum anisotropy and fitting the concentration where the difference between minimum and maximum is half. FIDA results in a distribution that is fitted with a gaussian, the properties of which give information about the complexes formed in the system. For each concen-

tration of RNAP the FIDA is measured and a Gaussian is fitted to determine the hydrodynamic radius. The change in hydrodynamic radius with increasing RNAP is then then fitted to find the dissociation constant. The FIDA data contains some large spikes, likely aggregates, which resulted in large errors and meant not all results could be fitted. Especially the low concentrations of RNAP had poor fits. This may explain the wide range in dissociation constants with FIDA as the low concentrations are important to determine the baseline for the fit of the dissociation constant.

The non-specific binding energy used in the model is 11.5 k_BT, based on a preliminary experiment on the effect of non-specific DNA on the reaction rate, see appendix D for details. This is equivalent to a dissociation constant of 10.1 µM. The non-specific dissociation constant was also measured with FA and FIDA to corroborate the results from the rate experiments. The measurements of non-specific sites are inconsistent and seem difficult to determine with FA. When the same FA data from figure 4 is fitted with two binding sites, there seems to be a plausible second binding site at around 1000 nM, but the fit is poor. Measurements with just nonspecific DNA yield an affinity of around 500 nM, similar to the specific affinity. With FIDA the non-specific affinity was measured to be around 1000 nM, which would be in line with the plausible second site measured with FA. However if we compare this to the binding energy found in the non-specific rate experiment, the dissociation constant is an order of magnitude smaller. This might be due to the nature of the FA and FIDA measurements, neither measure the degree of binding directly, instead relying on a related quantity to estimate the binding of RNAP. When RNAP binds to nonspecific DNA there are many possible statistical configurations in which the RNAP can bind, however there is only space for a certain number of RNAP. The maximum anisotropy is reached before all the possible binding sites are occupied by RNAP, due to excluded volume effects. The consequence of this is that the measured dissociation constant is much lower than the actual dissociation constant as only a fraction of the total possible number of non-specific binding sites can be occupied at the same time. In order to accurately measure the non-specific dissociation constant, the concentration of non-specific sites in the system should be an order of magnitude larger than the dissociation constant. Since this is necessarily not the case when titrating with RNAP, this method may give inaccurate results. The concentration of DNA during our bindings assays is far below the specific dissociation constant of 500 nM, this should also be below the non-specific dissociation constant, so it is very possible that these measurements are inaccurate. This effect can be circumvented by measuring the complex formation with a constant concentration of RNAP and instead titrating with DNA. If the resulting dissociation constant is much higher, that would be a strong indication that there are indeed excluded volume effects suppressing the dissociation constant of non-specific sites. There are other complications with measuring the FA in this manner, as the RNAP would have to be labeled to be able to measure the anisotropy.

5.3 Optimizing quantification of T7 RNAP assay

The simplest test of the model is to investigate the binding of RNAP to the DNA. The concentration of RNAP and DNA was varied over three orders of magnitude to find the most optimal reaction conditions for future experiments. For each point, equation 8 was solved for the fugacity λ_p with ϵ_{ps} set to 14.5 k_BT and ϵ_{pns} set to 11.5 k_BT as determined in the previous section. The fugacity was then inserted into equation 4 to find the occupation θ.

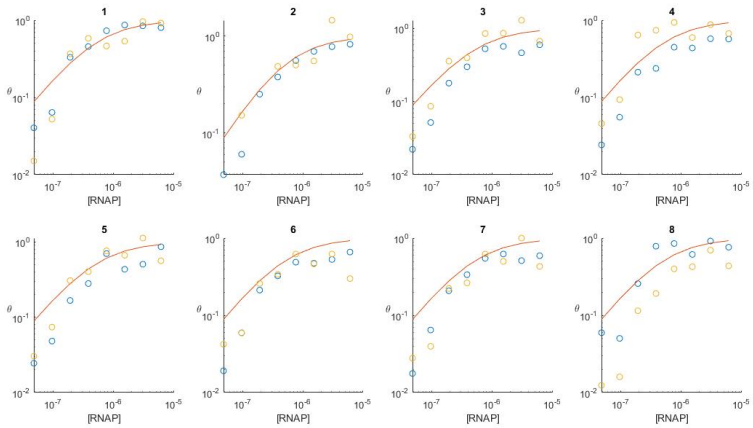

Figure 5: Activity scaled to theory as described in section 5.1 versus the concentration of RNAP in mol/L . Figure 1-8 correspond to a DNA concentration of 5.8 nM, 2.9 nM, 1.45 nM, 0.72 nM, 0.36 nM, 0.18 nM, 0.09 nM, 0.045 nM respectively. Experimental data is scaled to the occupation of the specific sites.

Over the whole range of DNA concentrations the model agrees well with the experiments see figure 5. At low concentrations of DNA the data is clearly noisier, but still agrees quite well with the model, with the exception that the model is consistently off at the lowest concentrations of RNAP. In figure 6 it is very clear that there is a shift between theory and experiment at the two lowest concentrations of RNAP.
The activity scales linearly with the amount of DNA, see figure 6. This is expected as the concentration of DNA is well below the K_d of 485 nM. RNAP is always in excess, meaning there is a significant portion of the RNAP dissolved in the water and when DNA is increased there is enough RNAP available to bind to the promoters. When increasing the concentration of DNA, the activity should eventually plateau when all of the available RNAP is bound to DNA and thus no free RNAP is available to bind to additional DNA.

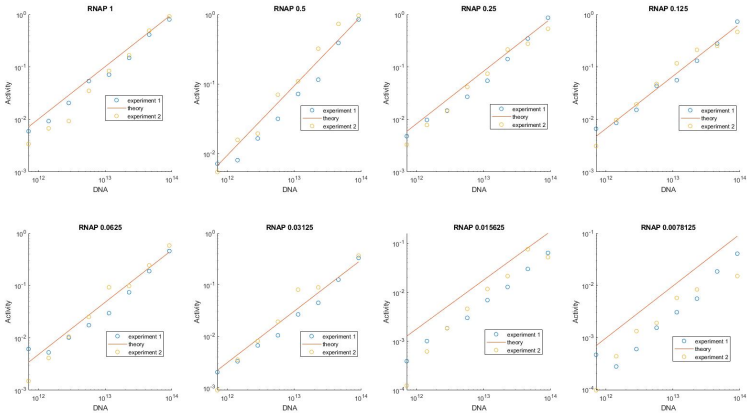

Figure 6: Activity scaled to theory as described in section 5.1 versus the concentration of DNA in mol/L. Figure 1-8 correspond to a RNAP concentration of 6.0 μM, 3.0 μM, 1.5 μM, 0.75 μM, 0.375 μM, 0.1875 μM, 0.09375 μM, 0.046875 μM respectively.

In figure 7 the data was normalized with respect to the DNA concentration. This normalized activity is compared to the theoretical occupation of the DNA. This causes the data of all the different DNA concentrations to collapse onto a single adsorption curve, see figure 7 and 8.

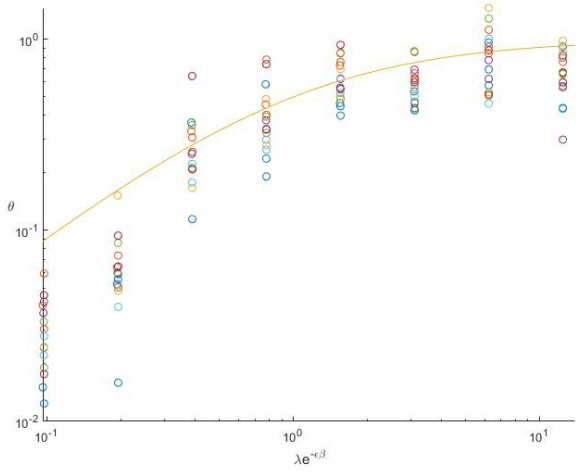

Figure 7: Activity scaled to occupation θ of the specific sites versus $\lambda e^{-\epsilon\beta}$, the statistical weight in mol/L of RNAP bound to the promoter. Statistical weight is calculated from the concentration of RNAP.

At low concentrations of RNAP there is a discrepancy between theory and experiment. The theory overestimates the actual transcription rate. This points to a system where less RNAP is active then expected. This is likely because the RNAP adsorbs onto the plastic surface of the plate. This results in a decrease in the fugacity as less RNAP is available to bind to the DNA. It should be

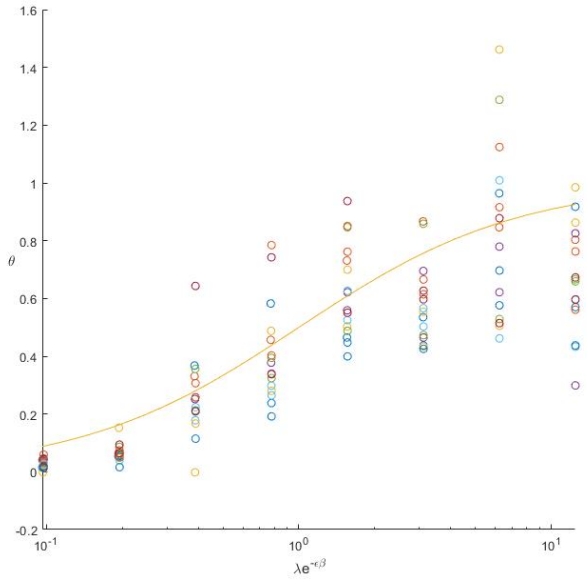

Figure 8: Lin log plot of the activity scaled to occupation θ of the specific sites versus $\lambda e^{-\epsilon\beta}$, the statistical weight in mol/L of RNAP bound to the promoter. Statistical weight is calculated from the concentration of RNAP in mol/L.

possible to construct the same adsorption curve by calculating the fugacity from the concentration of DNA. Additional experiments with higher concentrations of DNA could be conducted to confirm that the transcription rate follows a similar adsorption isotherm as a function of DNA.

5.4 Effect of BSA at low concentrations

In the following experiments various systems with BSA are compared to systems without BSA. In some in-vitro studies, BSA is added to the reaction to prevent enzymes from sticking to surfaces [14]. To test the hypothesis that RNAP adsorbs to the plastic surface, BSA was added. The measured activity was then compared to the activity without BSA. If the activity at low concentrations is higher when BSA is present, this is an indication that there is indeed adsorption of RNAP to the well surface.

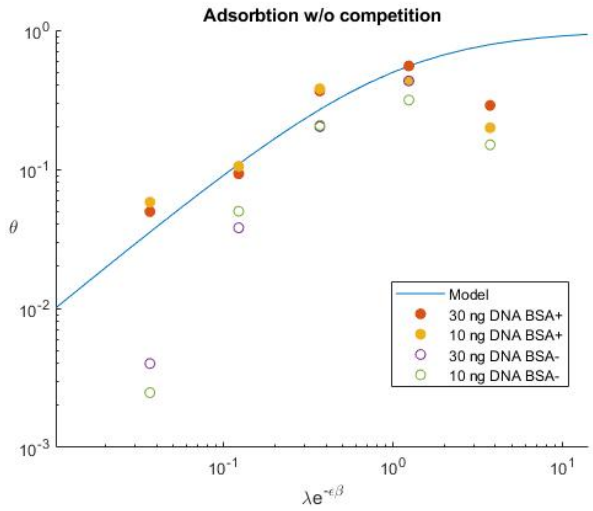

Figure 9: Activity scaled to occupation of the specific sites versus the statistical weight of RNAP bound to the promoter. Statistical weight is calculated from the concentration of RNAP in mol/L. BSA was added to experiments indicated with BSA+

In figure 9 was added the system with several concentrations of DNA and compared to the same system without BSA. The data with BSA corresponds much better to the theory. This indicates that there could indeed by competition for RNAP by the well surface. Adding BSA to the system occupies these binding spots and as a result there is more RNAP available to bind to the promoters. This in turn causes an increase in activity when BSA is present.

To account for competition for RNAP by the well surface an additional term was be added to the mass balance as per equation 9. The number of available binding spots was approximated by taking the dimensions of the plate well and calculating the available surface area. The available surface area was divided by the approximate surface area of an RNAP molecule, a detailed calculation can be found in appendix E. According to this estimate the competitive sites can bind around 68 nM of RNAP.

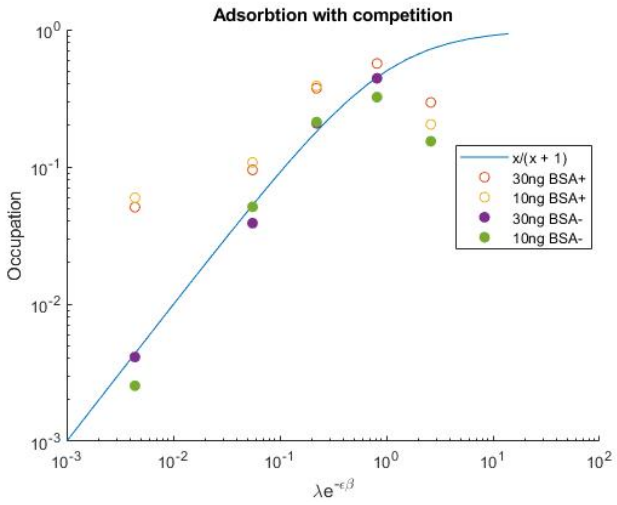

Figure 10: Activity scaled to occupation of the specific sites versus the statistical weight of RNAP bound to the promoter. Statistical weight is calculated from the concentration of RNAP in mol/L. BSA was added to experiments indicated with BSA+. Theory is plotted according the model with additional competitive sites with a competitive binding energy of 18.5 k_BT

In figure 10 the model was adapted to include the competition by the well surface. While this approximation likely overestimates the number of binding sites, it should still be in the order of magnitude of the maximum number of RNAP that can adsorb to the surface. The binding energy to the surface was estimated by first scaling the data with BSA to the activity as described in section 5.1 . The resulting scaling factor is then applied to the data without BSA. The scaled data was then compared to the model with competition with an initial competitive binding energy of 19.5 k_BT. This was varied by steps of 0.5 k_BT and the binding energy with the best fit to the experimental data was chosen as the standard. This best fit was 18.5 k_B which can be converted to a dissociation constant of 9.2 nM. The effect of the competitive sites decreases after around 100 nM RNAP, at this point most if not all of the competitive sites are already bound and as such any additional RNAP is able to bind to the specific sites immediately.

With the competitive term the model agrees with the data without BSA. This agrees with the hypothesis that RNAP adsorbs onto the well surface. To further confirm this hypothesis, an experiment could be set up to test the effect of available surface area on the transcription rate. This could perhaps be achieved by performing a similar experiment with varying volume at low concentrations of RNAP, keeping all other variables constant.

To investigate the consistency of experiments with BSA and low concentration activity, the experiment was repeated with additional lower concentrations. As seen in figure 11 the model agrees very well with the experimental data. The

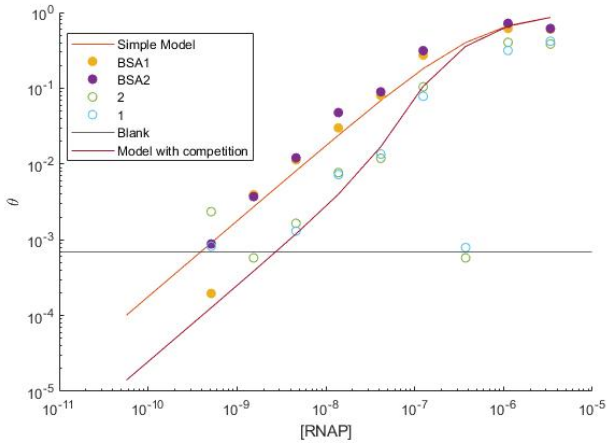

Figure 11: Activity scaled to occupation of the specific sites versus the concentration of RNAP in mol/L. BSA1 and BSA2: 7.7 nM T7 Broccoli DNA with 100 ng/µl BSA. 1 and 2: 7.7 nM T7 Broccoli DNA. The competitive model was calculated with a competitive binding energy of 18.5 k_BT

curve with competitions shows very nicely how the difference in activity is can be explained by RNAP binding to competitive sites. In appendix B slightly different binding energies were used to plot the theory. There is still a slight deviation at high concentrations of RNAP. The activity of the series without BSA is consistently slightly lower than with BSA, see appendix B.

One possible explanation for this is the aggregation of RNAP at high concentrations. BSA is known to prevent aggregation of proteins and is often used to that effect [9]. It is also used to improve the activity of T7 transcription assays [5]. Although RNAP doesn't typical aggregate at these concentrations [21], it is of a sufficiently high concentration in the stock solution of 150 µM. Perhaps BSA is able to help the RNAP dissolve, instead of remaining in kinetically stable aggregates formed in the stock solution. It is possible these aggregates are responsible for the spikes in the FIDA measurements discussed earlier, however the nature of the aggregates measured with FIDA is not known and could just as well be the DNA. An interesting addition to the model would be to add a term taking into account the aggregation of RNAP, which could give some indication of the plausibility of this hypothesis.

Another possible explanation is the self inhibition of RNAP at high concentrations. RNAP binding to the DNA shortly after one another can bump into each other either slowing down the lagging complex or completely displacing the leading complex. This induces instability in the transcription complex and in turn leads to a lower transcription rate [31]. In case of the latter BSA should not have any effect on the transcription at high concentrations, unless it has some stabilizing effect on the transcription complex.

6 Conclusion

We showed that Broccoli DNA can be used to quantitatively measure transcription rates in vitro and compared the experimental rates to a model based on statistical thermodynamics. RNA polymerase and the amount of DNA were varied and the corresponding change in transcription rate was observed. The experimental results were compared with the grand canonical model with which we calculate the occupation of the promoter sites. The K_d of the specific sites found with FA and FIDA are consistent between experiments and literature. From this we calculated a specific binding energy of 14.5 k_BT, which is consistent across rate experiments, model and independent binding assays. The K_d of the nonspecific sites measured is very close to that of the specific sites, this could due to volume exclusion effect as the concentration of DNA used in the binding experiments is very low. From rate experiments we predicted a non-specific binding energy of 11.5 k_BT, which is not consistent with the independent binding assays. To understand this better, the non-specific binding energy should be investigated further. To investigate the nonspecific binding energy, care should be taken to keep the concentration of DNA at least as high as the expected dissociation constant. More experiments should be conducted to confirm the binding energies and investigate the rate limiting step within the transcription process. This does not seem to have any significant consequences for the model, as we are able to fit the experimental data with the K_d obtained in the binding experiments. The model can be adapted to account for unexpected competition by plastic surfaces by adding a competitive term. This competition can be prevented by adding BSA to the system, simplifying the system. With this method and model we have a strong foundation to study transcription quantitatively. This will allow for systematic studies of gene regulation.

7 Outlook

To further confirm the specific dissociation constant it would be good to also perform a study of the complex formation with a constant concentration of RNAP and increasing concentration of DNA. Is is especially interesting to apply this to measure the nonspecific binding energy to investigate whether there are indeed excluded volume effects that suppress the apparent dissociation constant in our FA measurements. The reproducibility of the experiments presented in this work should be investigated as there is a lot of variation in results from one experiment to the next. With this new method it would be very interesting to investigate more complex gene regulation with for example lac repressor, an activator or other additional transcription factors. It could also be used to investigate the effect of adding additional non-specific DNA or perhaps additional competitive sites. For these experiments, moving to a system based on Ecoli instead of T7 would be advantageous, as Ecoli RNAP has a high affinity for DNA, which would be very useful when investigating the effect of DNA on transcription. In this case most of the RNAP would be situated near the DNA.

The model can then be simplified by removing the term accounting for RNAP dissolved in the water.

8 Acknowledgements

I would like to thank Willem Kegel for introducing me to this project, his supervision and the many fruitful discussions. Gert Folkers for his practical supervision, performing several experiments on BSA and in particular for performing many experiments on the binding affinity of T7 RNAP to DNA. Amanda van der Sijs for helping with an experiment with BSA. Roy van Stralen for helping with the experimental setup and assisting with the preliminary experiments. And finally, Jasper Landman for helping me with the model and the useful discussions on the theory of gene regulation. I would also like to thank, in general, the members of the FCC research group for providing a comfortable and engaging workplace, as well as the participants in the friday brainstorms for the interesting discussions and input.

8.1 Specific experimental contributions

The optimization results in chapter 5.3 are a combination of data from two separate repetitions. The first optimization experiment of T7 RNAP assay was performed with the assistance of Roy van Stralen. Represented in figures 5-8. The second optimization experiment of T7 RNAP assay, including calibration lines was performed by Gert Folkers. Represented in figures 5-8 and in appendix F, figures 35 and 36. The binding experiments discussed in chapter 5.2 were performed by Gert Folkers. Represented in figure 4. The preliminary experiment with LacI and non-specific DNA in appendix C was performed by Gert Folkers. Represented in appendix C, figure 25. The BSA experiment with additional lower concentrations in chapter 5.4 was performed by Amanda van der Sijs. Represented in figure 11, as well as in appendix D and E, figures 26-34.

References

[1] G. K. Ackers, A. D. Johnson, and M. A. Shea. "Quantitative model for gene regulation by λ phage repressor". In: *Proceedings of the National Academy of Sciences of the United States of America* 79.4 I (1982), pp. 1129–1133. ISSN: 00278424. DOI: 10.1073/pnas.79.4.1129.

[2] Khalid K. Alam et al. "A Fluorescent Split Aptamer for Visualizing RNA-RNA Assembly in Vivo". In: *ACS Synthetic Biology* 6.9 (Sept. 2017), pp. 1710–1721. ISSN: 21615063. DOI: 10.1021/acssynbio.7b00059. URL: https://pubs.acs.org/sharingguidelines.

[3] Rajiv P Bandwar et al. “Kinetic and thermodynamic basis of promoter strength: Multiple steps of transcription initiation by T7 RNA polymerase are modulated by the promoter sequence”. In: *Biochemistry* 41.11 (2002), pp. 3586–3595. ISSN: 00062960. DOI: 10.1021/bi0158472. URL: https://pubs.acs.org/sharingguidelines.

[4] Lacramioara Bintu et al. “Transcriptional regulation by the numbers: Models”. In: *Current Opinion in Genetics and Development* 15.2 (2005), pp. 116–124. ISSN: 0959437X. DOI: 10.1016/j.gde.2005.02.007.

[5] M. Chamberlin and J. Ring. “Characterization of T7-specific ribonucleic acid polymerase. 1. General properties of the enzymatic reaction and the template specificity of the enzyme.” In: *Journal of Biological Chemistry* 248.6 (1973), pp. 2235–2244. ISSN: 00219258. DOI: 10.1016/S0021-9258(19)44211-7.

[6] Graham MT Cheetham and Thomas A Steitz. “Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases”. In: *Current Opinion in Structural Biology* 10.1 (2000), pp. 117–123. ISSN: 0959-440X. DOI: https://doi.org/10.1016/S0959-440X(99)00058-5. URL: https://www.sciencedirect.com/science/article/pii/S0959440X99000585.

[7] Fabio Chizzolini et al. “Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength Terms of Use”. In: *ACS Synth. Biol* 3 (2014), p. 371. DOI: 10.1021/sb4000977. URL: https://pubs.acs.org/sharingguidelines.

[8] Grigory S. Filonov et al. “Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution”. In: *Journal of the American Chemical Society* 136.46 (2014), pp. 16299–16308. ISSN: 15205126. DOI: 10.1021/ja508478x.

[9] Thomas E. Finn et al. “Serum albumin prevents protein aggregation and amyloid formation and retains chaperone-like activity in the presence of physiological ligands”. In: *Journal of Biological Chemistry* 287.25 (2012), pp. 21530–21540. ISSN: 00219258. DOI: 10.1074/jbc.M112.372961. URL: http://dx.doi.org/10.1074/jbc.M112.372961.

[10] Hernan G Garcia and Rob Phillips. “Quantitative dissection of the simple repression input-output function”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.29 (2011), pp. 12173–12178. ISSN: 00278424. DOI: 10.1073/pnas.1015616108.

[11] Richard A Ikeda and Charles C Richardson. “Interactions of the RNA polymerase of bacteriophage T7 with its promoter during binding and initiation of transcription”. In: *Proceedings of the National Academy of Sciences of the United States of America* 83.11 (1986), pp. 3614–3618. ISSN: 00278424. DOI: 10.1073/pnas.83.11.3614.

[12] Henrik Jensen and Jesper Østergaard. "Flow Induced Dispersion Analysis Quantifies Noncovalent Interactions in Nanoliter Samples". In: *Journal of the American Chemical Society* 132.12 (2010), pp. 4070–4071. DOI: 10.1021/ja100484d. URL: http://pubs.acs.org..

[13] Yiping Jia, Amarendra Kumar, and Smita S. Patel. "Equilibrium and stopped-flow kinetic studies of interaction between T7 RNA polymerase and its promoters measured by protein and 2-aminopurine fluorescence changes". In: *Journal of Biological Chemistry* 271.48 (1996), pp. 30451–30458. ISSN: 00219258. DOI: 10.1074/jbc.271.48.30451. URL: http://dx.doi.org/10.1074/jbc.271.48.30451.

[14] Zachary J. Kartje et al. "Revisiting T7 RNA polymerase transcription in vitro with the Broccoli RNA aptamer as a simplified real-time fluorescent reporter". In: *Journal of Biological Chemistry* 296 (Jan. 2021). ISSN: 1083351X. DOI: 10.1074/jbc.RA120.014553.

[15] Jasper Landman et al. "Self-consistent theory of transcriptional control in complex regulatory architectures". In: *PLoS ONE* 12.7 (July 2017), e0179235. ISSN: 19326203. DOI: 10.1371/journal.pone.0179235. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0179235.

[16] Xing Li et al. "Fluorophore-Promoted RNA Folding and Photostability Enables Imaging of Single Broccoli-Tagged mRNAs in Live Mammalian Cells". In: *Angewandte Chemie* 132.11 (Mar. 2020), pp. 4541–4548. ISSN: 0044-8249. DOI: 10.1002/ange.201914576. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ange.201914576.

[17] Jo L. Linpinsel and Graeme L. Conn. "General protocols for preparation of plasmid DNA template, RNA in vitro transcription, and RNA purification by denaturing PAGE". In: *Methods in Molecular Biology* 941 (2012), pp. 43–58. ISSN: 10643745. DOI: 10.1007/978-1-62703-113-4_4. URL: https://link-springer-com.proxy.library.uu.nl/protocol/10.1007/978-1-62703-113-4%7B%5C_%7D4.

[18] Cuihua Liu and Craig T Martin. "Promoter clearance by T7 RNA polymerase. Initial bubble collapse and transcript dissociation monitored by base analog fluorescence". In: *Journal of Biological Chemistry* 277.4 (2002), pp. 2725–2731. ISSN: 00219258. DOI: 10.1074/jbc.M108856200. URL: http://www.jbc.org.

[19] R Losick. "In Vitro Transcription". In: *Annual Review of Biochemistry* 41.1 (1972). PMID: 4628438, pp. 409–446. DOI: 10.1146/annurev.bi.41.070172.002205. URL: www.annualreviews.org.

[20] J R Lundblad, M Laurance, and R H Goodman. "Fluorescence polarization analysis of protein-DNA and protein-protein interactions." In: *Molecular Endocrinology* 10.6 (June 1996), pp. 607–612. ISSN: 0888-8809. DOI: 10.1210/mend.10.6.8776720. URL: https://academic.oup.com/mend/article-lookup/doi/10.1210/mend.10.6.8776720.

[21] Maribeth Maslak and Craig T. Martin. "Effects of Solution Conditions on the Steady-State Kinetics of Initiation of Transcription by T7 RNA Polymerase". In: *Biochemistry* 33.22 (1994), pp. 6918–6924. ISSN: 15204995. DOI: 10.1021/bi00188a022.

[22] Muir Morrison, Manuel Razo-Mejia, and Rob Phillips. "Reconciling kinetic and thermodynamic models of bacterial transcription". In: *PLoS Computational Biology* 17.1 (Jan. 2021), e1008572. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1008572. URL: https://doi.org/10.1371/journal.pcbi.1008572.

[23] Voraphoj Nilaratanakul, Debra A. Hauer, and Diane E. Griffin. "Development of encoded Broccoli RNA aptamers for live cell imaging of alphavirus genomic and subgenomic RNAs". In: *Scientific Reports* 10.1 (Dec. 2020), pp. 1–13. ISSN: 20452322. DOI: 10.1038/s41598-020-61573-3. URL: https://doi.org/10.1038/s41598-020-61573-3.

[24] Morten E. Pedersen, Jesper Østergaard, and Henrik Jensen. "Flow-induced dispersion analysis (FIDA) for protein quantification and characterization". In: *Methods in Molecular Biology*. Vol. 1972. Humana Press Inc., 2019, pp. 109–123. DOI: 10.1007/978-1-4939-9213-3_8. URL: https://doi.org/10.1007/978-1-4939-9213-3%7B%5C_%7D8,.

[25] Madeline A Shea and Gary K. Ackers. "The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation". In: *Journal of Molecular Biology* 181.2 (1985), pp. 211–230. ISSN: 00222836. DOI: 10.1016/0022-2836(85)90086-5.

[26] Gary M Skinner et al. "Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase: A single-molecule view of the transcription cycle". In: *Journal of Biological Chemistry* 279.5 (2004), pp. 3239–3244. ISSN: 00219258. DOI: 10.1074/jbc.M310471200. URL: http://www.jbc.org.

[27] R. Sousa. "T7 RNA Polymerase". In: *Encyclopedia of Biological Chemistry: Second Edition* (2013), pp. 355–359. ISSN: 0079-6603. DOI: 10.1016/B978-0-12-378630-2.00267-X.

[28] V. L. Tunitskaya and S. N. Kochetkov. "Structural-functional analysis of bacteriophage T7 RNA polymerase". In: *Biochemistry (Moscow)* 67.10 (2002), pp. 1124–1135. ISSN: 00062979. DOI: 10.1023/A:1020911223250.

[29] Andrea Újvári and Craig T. Martin. "Thermodynamic and kinetic measurements of promoter binding by T7 RNA polymerase". In: *Biochemistry* 35.46 (Nov. 1996), pp. 14574–14582. ISSN: 00062960. DOI: 10.1021/bi961165g. URL: https://pubs.acs.org/sharingguidelines.

[30] Franz M. Weinert et al. "Scaling of gene expression with transcription-factor fugacity". In: *Physical Review Letters* 113.25 (2014), pp. 1–5. ISSN: 10797114. DOI: 10.1103/PhysRevLett.113.258101.

[31] Yi Zhou and Craig T Martin. “Observed instability of T7 RNA polymerase elongation complexes can be dominated by collision-induced ”bumping””. In: *Journal of Biological Chemistry* 281.34 (2006), pp. 24441–24448. ISSN: 00219258. DOI: 10.1074/jbc.M604369200.

9 Appendix

A Finding reaction rate by linear fit

To find the initial reaction rate at a certain concentration of RNAP and DNA, the fluorescence versus the time is fitted from 3 to 30 minutes, see figure 12. The first few minutes are cut off to allow the temperature to stabilize and reaction to reach it's maximum rate. Data after 30 minutes tends to be subject to depletion of nucleotides and a subsequent decrease in reaction rate, so that is also omitted.

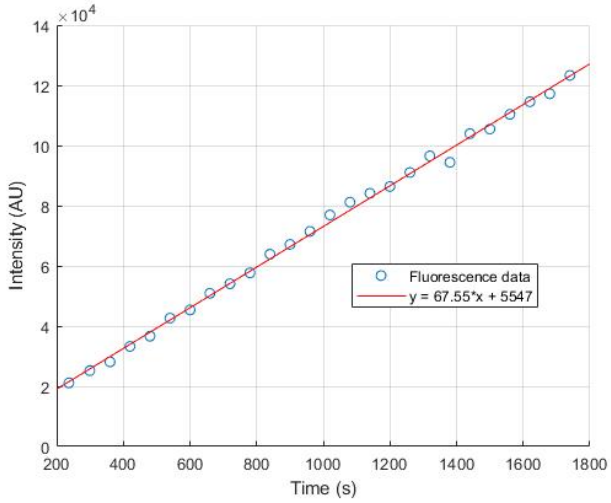

Figure 12: Linear fit of the first 3-30 minutes of a standard transcription reaction. Reaction carried out with 17.6 ng of T7 Broccoli DNA and 125 nM RNAP.

B Additional experiments with BSA

In figure 13 it is clear that the model does not agree with experiment. when the binding affinity is increased by 2 k_BT, see figure 14 the model does fit the data. In figure 15 and 16 two lower concentrations of DNA are very noisy, but do still fit the higher binding affinity model. This is possibly because a different type of DNA was used, the T7-O2-Brocc DNA has an extra lac operon near the promoter site. This slight difference in promoter environment could be the cause of the difference in binding energy. There is also a difference in sequence length, but this was taken into account by the model and does not have a large effect on the occupation of specific sites.

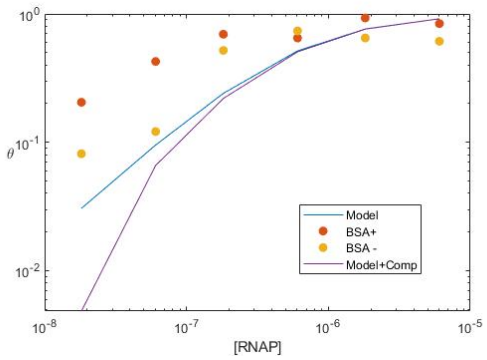

Figure 13: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA+ 25 ng T7-O2 Broccoli DNA with 100 ng/ul BSA. BSA- 25 ng T7-O2 Broccoli DNA. Theory plotted with a specific binding affinity of 14.5 $k_B T$

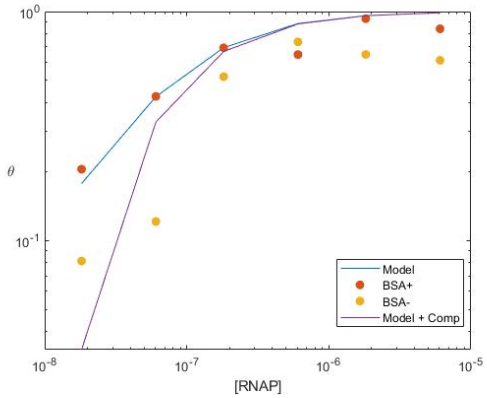

Figure 14: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA+ 25 ng T7-O2 Broccoli DNA with 100 ng/ul BSA. BSA- 25 ng T7-O2 Broccoli DNA. Theory plotted with a specific binding affinity of 16.5 $k_B T$

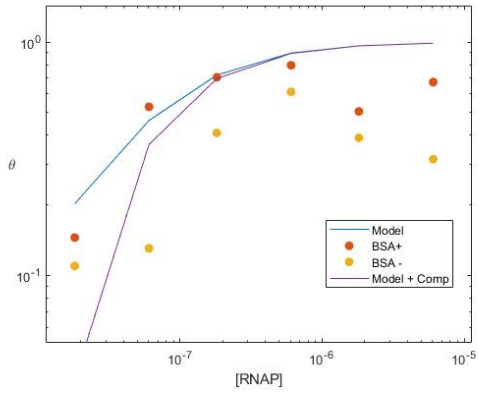

Figure 15: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA+ 7.5 ng T7-O2 Broccoli DNA with 100 ng/ul BSA. BSA- 7.5 ng T7-O2 Broccoli DNA. Theory plotted with a specific binding affinity of 16.5 $k_B T$

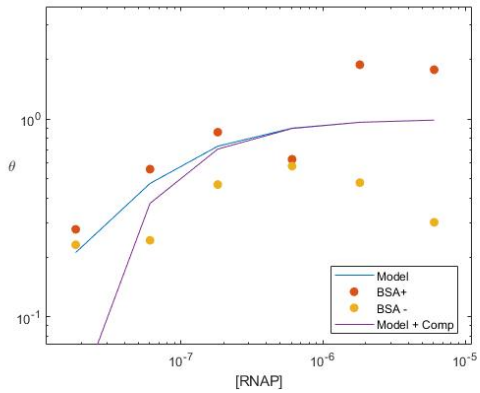

Figure 16: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA+ 2.5 ng T7-O2 Broccoli DNA with 100 ng/ul BSA. BSA- 2.5 ng T7-O2 Broccoli DNA. Theory plotted with a specific binding affinity of 16.5 $k_B T$

To compare the following experiments were performed with T7-Brocc DNA, similar to the T7-O1 sequence used in other experiments.

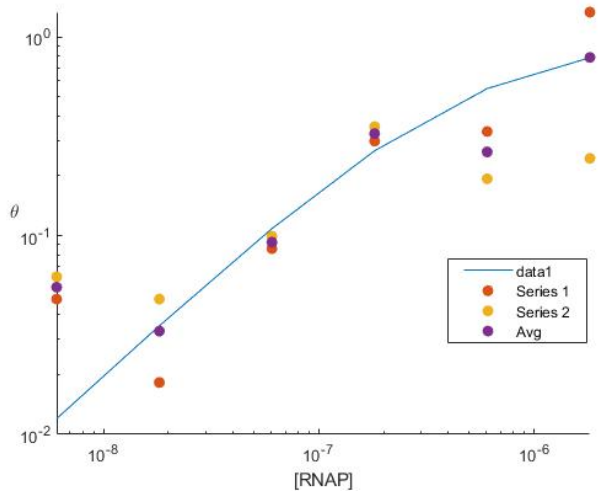

Figure 17: Activity scaled to occupation of the specific sites versus the concentration of RNAP. Series 1 and 2: 30 ng T7 Broccoli DNA with 100 ng/ul BSA. Avg: average of series 1 and 2. Theory was calculated with a specific binding affinity of 14.5 k_BT

In this experiment the binding energy is indeed 2 k_BT lower, the same as similar experiments before. However in this series, at lower RNAP concentrations the activity is much higher than expected. This phenomenon becomes more pronounced at low concentrations of DNA.

To investigate this phenomenon, an experiment was performed with much lower concentrations of RNAP to see when the activity could no longer be measured. The same effect as in figure 17, 18 and 19 is observed again in figure 20 The blank value was plotted as a horizontal line to illustrate that the activities measured are not just noise and orders of magnitude larger than the blank. The activity at low concentration does not show a linear response with respect to RNAP. This is very strange as with any system, when the concentration of RNAP is very low, the response should always be linear as there are more than enough unoccupied promoter sites to bind to. Strangely there is still significant activity even when the activity should fall below the blank according to the model.

In figure 21 the experiment was repeated with a higher concentration of DNA to check whether this phenomenon has something to do with a low concentration of DNA. In this case the lower concentrations do behave as expected, but there is some deviation at higher concentrations and the data with BSA does not significantly differ from the data without. It is possible some BSA was accidentally added during this experiment. It was observed that the NTP

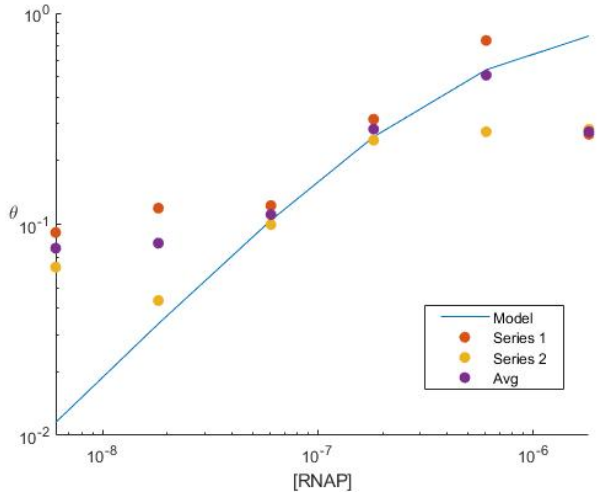

Figure 18: Activity scaled to occupation of the specific sites versus the concentration of RNAP. Series 1 and 2: 10 ng T7 Broccoli DNA with 100 ng/ul BSA. Avg: average of Series 1 and 2. Theory was calculated with a specific binding affinity of 14.5 k_BT

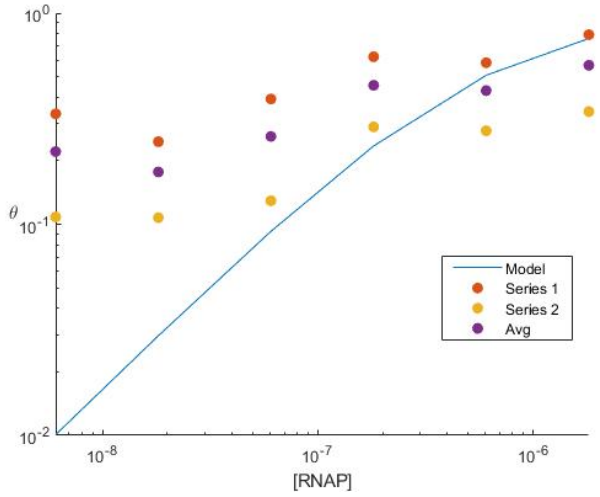

Figure 19: Activity scaled to occupation of the specific sites versus the concentration of RNAP. Series 1 and 2: 10 ng T7 Broccoli DNA with 100 ng/ul BSA. Avg: average of Series 1 and 2. Theory was calculated with a specific binding affinity of 14.5 k_BT

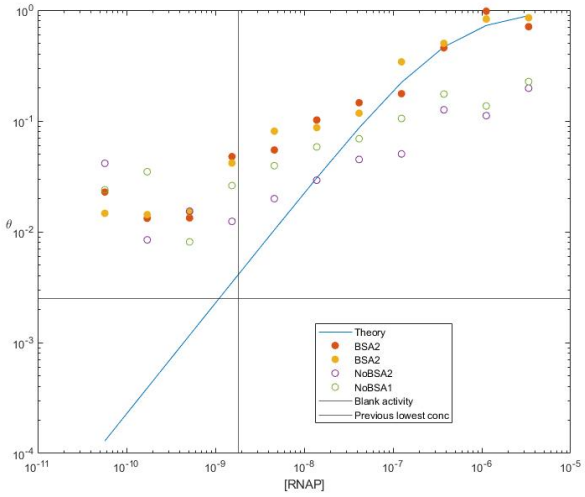

Figure 20: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. noBSA1 and noBSA2: 17.6 ng T7 Broccoli DNA. Model was calculated with a specific binding affinity of 14.5 k_BT.

used to initiate the reaction was very viscous, which could point to presence of protein.

The experiment was repeated again with just the lower concentrations, see figure 22. In this experiment the low concentrations behave as expected both with and without BSA. In figure 23 the data is plotted together with figure 20 and in in figure 24 it is plotted together with figure 21

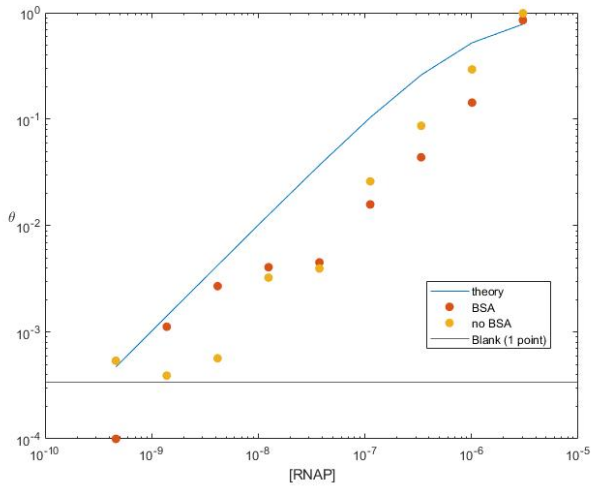

Figure 21: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA: 44 ng T7 Broccoli DNA with 100 ng/ul BSA. no BSA: 44 ng T7 Broccoli DNA. Model was calculated with a specific binding affinity of 14.5 k_BT.

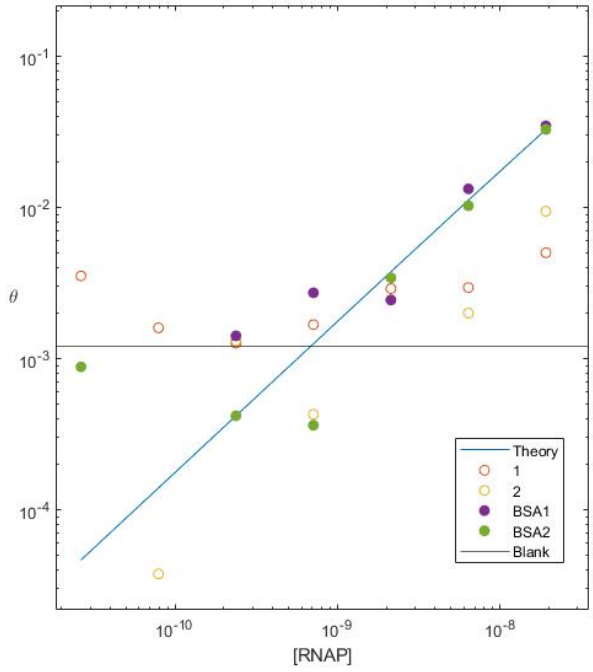

Figure 22: Activity scaled to occupation of the specific sites versus the concentration of RNAP. 1 and 2: 17.6 ng T7 Broccoli DNA. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. Model was calculated with a specific binding energy of 14.5 k_BT.

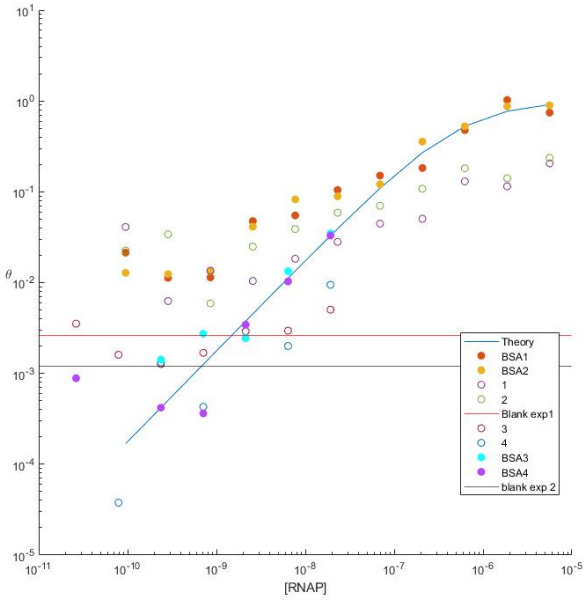

Figure 23: Combination of 20 and 22.1-4 17.6 ng T7 Broccoli DNA. BSA1-BSA4 1-4 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA

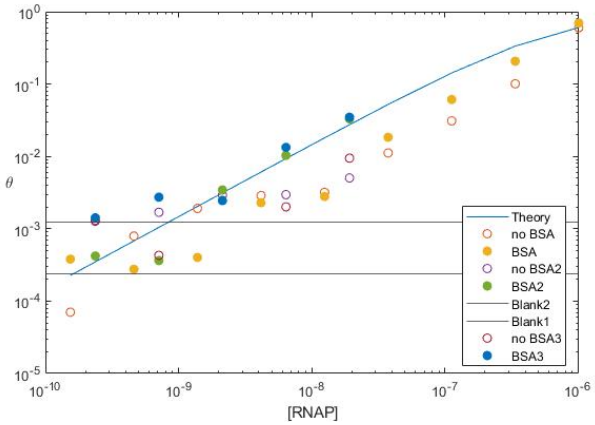

Figure 24: Combination of 21 and 22. 1-2 44 ng T7 Broccoli DNA. 3-4 17.6 ng T7 Broccoli DNA

. BSA1-2 44 ng T7 Broccoli DNA with BSA. BSA3-4 17.6 ng T7 Broccoli DNA.

C Repression by LacI

To take the repressor into account we need to add an additional state to the partition function, where the site is bound by a repressor. $\Xi_s = (1 + \lambda_p e^{-\beta\epsilon_p} + \lambda_r e^{-\beta\epsilon_s})^{N_s}$ The fugacity of the repressor is given by :$\lambda_r = e^{\beta\mu_R}$. This new partition function is applied to equation 3 which yields:

$< P_s >= \lambda_p \frac{\partial ln\Xi_s}{\partial\lambda_s} = N_s \frac{\lambda_p e^{-\beta\epsilon_p}}{1+\lambda_p e^{-\beta\epsilon_p}+\lambda_r e^{-\beta\epsilon_s}} = N_s\theta$ where θ is the occupation of the binding site $\theta = \frac{\lambda_p e^{-\beta\epsilon_p}}{1+\lambda_p e^{-\beta\epsilon_p}+\lambda_r e^{-\beta\epsilon_s}}$ Applying this to the mass balance:

$$[P] = [N_s]\frac{\lambda_p e^{-\beta\epsilon_{ps}}}{1+\lambda_p e^{-\beta\epsilon_{ps}}+\lambda_r e^{-\beta\epsilon_{rs}}} + [N_{ns}]\frac{\lambda_{pns} e^{-\beta\epsilon_{pns}}}{1+\lambda_p e^{-\beta\epsilon_{pns}}+\lambda_r e^{-\beta\epsilon_{rns}}} + [N_c]\frac{\lambda_p e^{-\beta\epsilon_{pc}}}{1+\lambda_p e^{-\beta\epsilon_{pc}}+\lambda_r e^{-\beta\epsilon_{rc}}} + \lambda_p$$

There are now two unknowns in this equation; λ_p and λ_r. We need to construct a similar mass balance for the number of repressors denoted by R.

$R =< R_s > + < R_{ns} > + < R_c > + < R_{aq} >$ Following the same steps yields:

$$[R] = [N_s]\frac{\lambda_r e^{-\beta\epsilon_{rs}}}{1+\lambda_p e^{-\beta\epsilon_{ps}}+\lambda_r e^{-\beta\epsilon_{rs}}} + [N_{ns}]\frac{\lambda_{rns} e^{-\beta\epsilon_{rns}}}{1+\lambda_p e^{-\beta\epsilon_{pns}}+\lambda_r e^{-\beta\epsilon_{rns}}} + [N_c]\frac{\lambda_r e^{-\beta\epsilon_{rc}}}{1+\lambda_p e^{-\beta\epsilon_{pc}}+\lambda_r e^{-\beta\epsilon_{rc}}} + \lambda_r$$

In this case the last term is the effective concentration of repressor instead of RNAP. With two equations for two unknowns we can calculate both λ_p and λ_r by solving it as a system of equations.

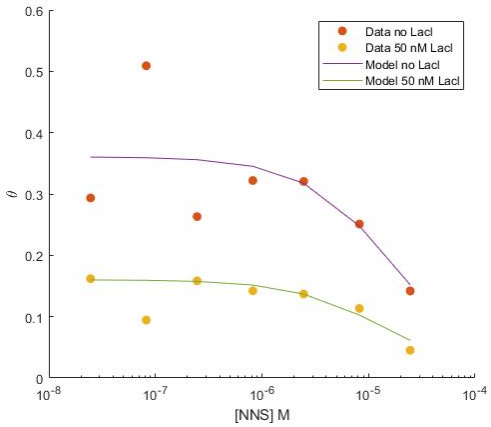

Figure 25: Activity scaled to occupation of the specific sites versus the molar concentration of Non specific sites [Nns]. Comparison of a system with 3.6 nM T7-O2 Broccoli DNA and 300 nM RNAP with and without 50 nM LacI repressor. Theory plotted with a specific binding energy of 14.5 k_BT and non-specific binding energy of 11.5 k_BT

An example of a system with repression can be seen in figure 25. This figure shows the effect of non-specific DNA on the activity of the RNAP, with

and without LacI repressor present. From this data set we can estimate the non-specific binding energy by fitting the data, results in a non-specific binding energy of around 11.5 k_BT. This values was used for most of the calculations performed in this work. Note that this is just a single experiment, but the important part is that the non-specific binding energy is several k_BT lower then the specific binding of 14.5 k_BT. This fit assumes that every base in the non-specific DNA is a potential binding site, which means this estimate of the number of non-specific binding sites likely overestimates the number of sites. This binding energy is consistent with the model as the number of non-specific sites is always estimated in the same manner for all experiments. The actual binding energy of non-specific sites could be a bit higher since there could be less actual bindings sites on the DNA.

D Precision and Accuracy of Binding Energy

The experimental results with BSA were modeled with several different specific binding energies form 12.5 k_BT to 16.5 k_BT, to get a grip on how well the binding energy measured in the FA experiments fits with the rate experiments. In the following figures the correlation coefficient R is included in the title.The specific binding energy of 14.5 k_BT produces the best fitting model, which is in line with the FA measurements. When the binding energy is changed by 1 k_BT, the decrease in correlation coefficient is not so significant, but still present. When the binding energy is shifted by 2 k_BT there is a strong effect on the correlation coefficient and it clear by eye that the model no longer fits as well.

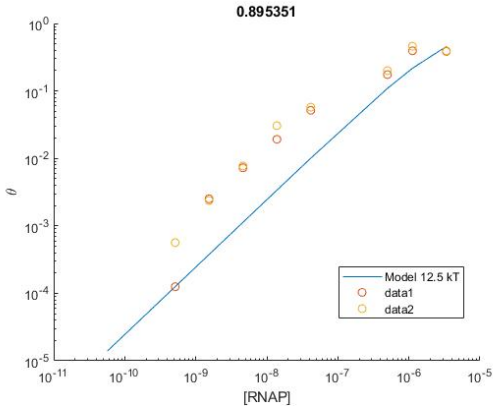

Figure 26: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. Model was calculated with a specific binding affinity of 12.5 k_BT.

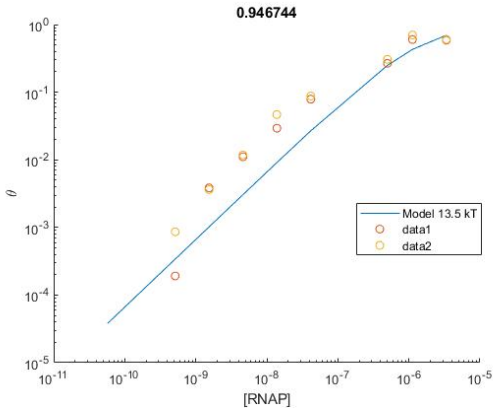

Figure 27: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. Model was calculated with a specific binding affinity of 13.5 k_BT.

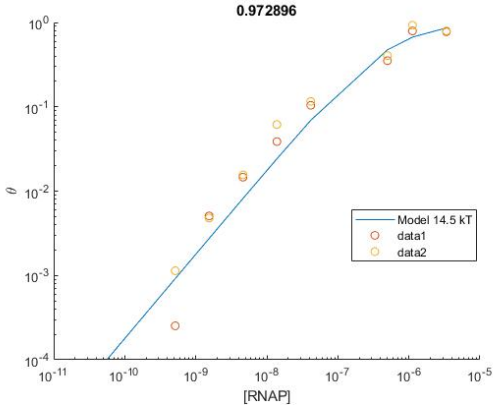

Figure 28: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. Model was calculated with a specific binding affinity of 14.5 k_BT.

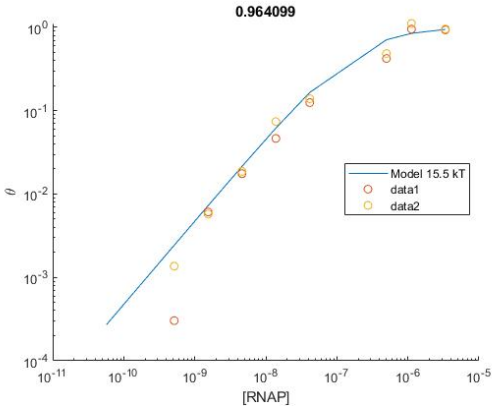

Figure 29: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. Model was calculated with a specific binding affinity of 15.5 k_BT.

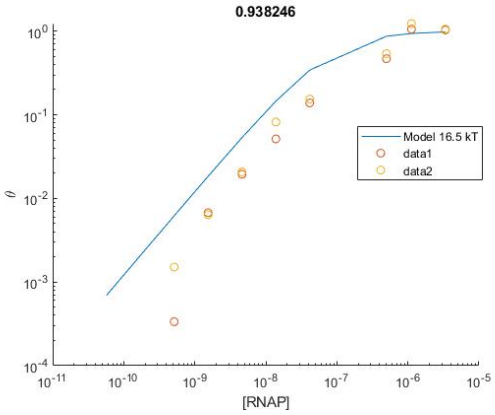

Figure 30: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. Model was calculated with a specific binding affinity of 16.5 k_BT.

E Estimating the number of adsorption sites in the well

To calculate the effect of competition by adsorption of RNAP to the surface of the reaction well, it is necessary to make an estimate of the number of RNAP that can adsorb to surface. The available surface area of the well was calculated with the surface of a truncated cone with a height of 4.7 mm, upper radius of 3.3 mm and lower radius of 1.84 mm. The upper radius of the cone is excluded from the surface calculation. π The available surface area is then $4.10 * 10^{-5}$. The projected area of adsorption for the RNAP was calculated by assuming the RNAP is spherical and directly related to the molecular weight of the protein according to the following equation: $M = \rho\frac{4}{3}\pi r^3$ Where ρ is the density of the protein, which is assumed to be of the same order of magnitude as water or 1 gcm^{-3}.
With a mass of around 99 kDa or 99000 g/mole, this results in a radius of 3.40 nm. The projected area of the RNAP is then $\pi * r^2 = 3.63 * 10^{-17} m^2$ This calculation likely underestimates the size of the RNAP as it does not take into account any form of hydration layer or possible shape change as a result of folding which could occur during the binding to the surface. The available surface area is divided by the projected area of the RNAP with maximum packing of circles to estimate the number of binding sites which gives $1.03 * 10^{12}$ sites. The number of sites where RNAP can be bound is equivalent to a concentration of 68 nM, to give an idea of how much RNAP can be bound. This is a very rough estimate that disregards any interactions between the RNAP and assumes binding events are independent, but should still give an idea of the order of magnitude of bound RNAP. Figure 31 shows the model with 68 nM competitive sites, with a competitive binding energy of 18.5 k_BT the model fits the data very well. In figures 33, **??** and 34 the same model was plotted with 29 nM competitive sites to give an idea of how a lower estimate would influence the model. In this case the binding energy has to be adjusted to 19.5 k_BT. The model with 68 nM of competitive sites looks to fit the data best, as here the decrease around 100 nM RNAP is clearly represented. With 29 nM sites the same decrease is much less pronounced and also overestimates the activity around 30 nM.

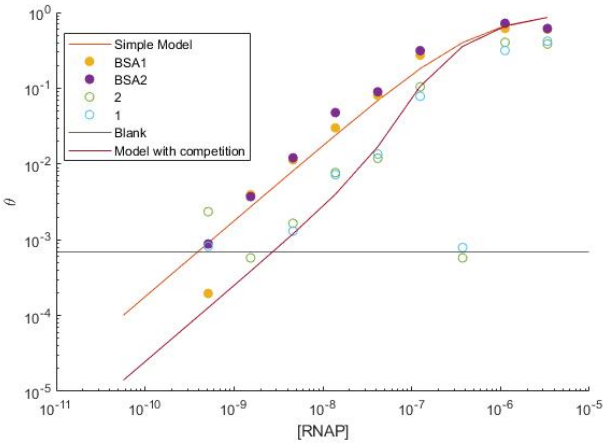

Figure 31: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. noBSA1 and noBSA2: 17.6 ng T7 Broccoli DNA. Model was calculated with a binding affinity of 14.5 k_BT. Competitive model was calculated with a specific binding affinity of 14.5 k_BT and a competitive binding energy of 18.5 k_BT. Concentration of competitive sites was set to 68 nM

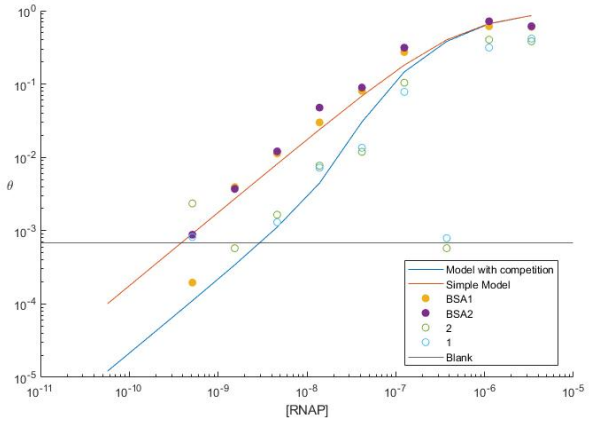

Figure 32: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. noBSA1 and noBSA2: 17.6 ng T7 Broccoli DNA. Model was calculated with a binding affinity of 14.5 k_BT. Competitive model was calculated with a specific binding affinity of 14.5 k_BT and a competitive binding energy of 19.5 k_BT. Concentration of competitive sites was set to 29 nM.

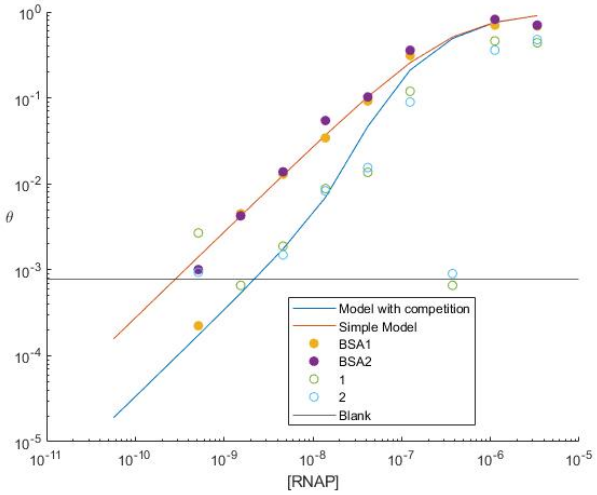

Figure 33: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. noBSA1 and noBSA2: 17.6 ng T7 Broccoli DNA. Model was calculated with a binding affinity of 15 k_BT. Competitive model was calculated with a specific binding affinity of 15 k_BT and a competitive binding energy of 19.5 k_BT. Concentration of competitive sites was set to 29 nM.

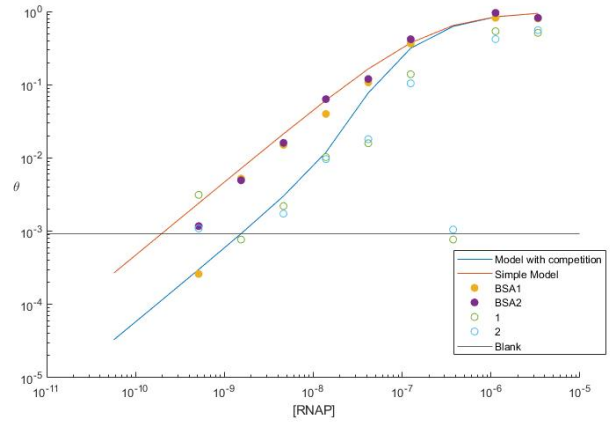

Figure 34: Activity scaled to occupation of the specific sites versus the concentration of RNAP. BSA1 and BSA2: 17.6 ng T7 Broccoli DNA with 100 ng/ul BSA. noBSA1 and noBSA2: 17.6 ng T7 Broccoli DNA. Model was calculated with a binding affinity of 15.5 k_BT. Competitive model was calculated with a specific binding affinity of 15.5 k_BT and a competitive binding energy of 19.5 k_BT. Concentration of competitive sites was set to 29 nM.

F Calibration line for Broccoli mRNA

A calibration line was made ranging from 8.8 ng to 8.2 µg and measured on both the Fluostar Optima (figure 35) and the SpectraMax iD3 Multi-Mode Microplate Reader (figure 36).
The relation between fluorescence (y) and amount of Broccoli mRNA (x) was found by linear fit.
For the Fluostar Optima:

$$y = 10380x - 199.11$$

For the SpectraMax iD3 Multi-Mode Microplate Reader:

$$y = 3.8854 * 10^5 x + 21256$$

Note that these calibration lines should only be used for transcription rates found in the same experiment as there are many settings and reaction conditions that can vary from experiment to experiment. Should the transcription rate in terms of Ms^{-1} be of interest, a calibration line can be included in the experiment.

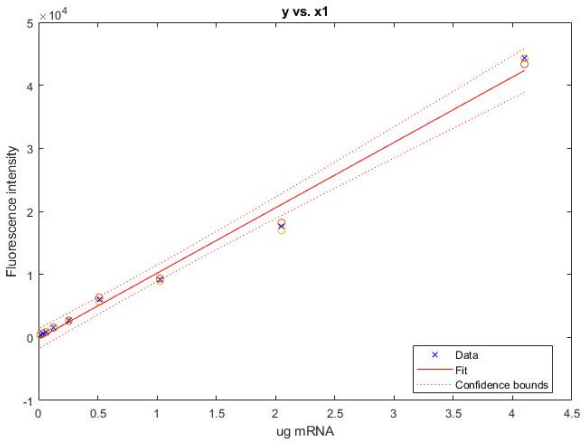

Figure 35: Calibration line for Broccoli mRNA on the Fluostar Optima. Known value of Broccoli mRNA on the x axis versus fluorescence intensity on the y axis.

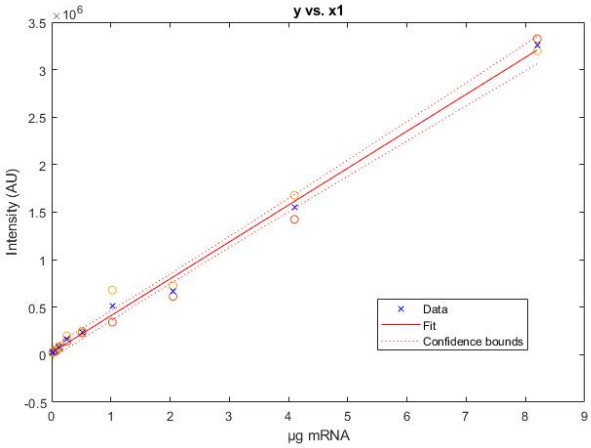

Figure 36: Calibration line for Broccoli mRNA on the SpectraMax iD3 Multi-Mode Microplate Reader. Known value of Broccoli mRNA on the x axis versus fluorescence intensity on the y axis.