POLITECNICO DI TORINO

Master's Degree in Physics of Complex Systems

Master's Degree Thesis

A Non-equilibrium model of the ParBS bacterial system

Supervisors

Prof. Andrea PAGNANI

Prof. Paolo DE LOS RIOS

Candidate

Andrea SALATI

October 2022

Abstract

ParABS systems are a tripartite mechanism that ensure chromosome and low-copy-number plasmid separation in most bacterial species. This thesis focuses on the interaction of two of these elements: centromere DNA sequence ParS and the DNA-binding protein ParB.

A new analytical coarse-grained model for the diffusion of ParB dimers on DNA is proposed. A first diffusive model was introduced in [1] but it ignores both the structural transitions of the dimer and the change of nucleotide it is bound to (either CDP or CTP). This model distinguishes between the open and closed states of the protein, therefore taking also in account the transition between these two conformations. Adding this feature in the model is crucial for a more faithful description of the ParBS interaction. Indeed experiments show that the ParS sequence catalyzes the closing transition of ParB, creating clusters of them around itself. The density distribution of ParB over DNA is therefore very peaked near the centromere. The main goal of this thesis is to recreate such distribution through a biophysical model.

From a thermodynamical point of view the equilibrium distribution is completely flat. The system is therefore considered to be in a Non Equilibrium Steady State (NESS). The only way this NESS can maintain itself is through the hydrolysis of CTP. In this thesis a continuous limit of the previously (numerically solved) discrete lattice model has been performed in order to obtain an approximate but analytical solution. The latter has been compared with the equilibrium one, gaining information on how much energy this system dissipates.

Table of Contents

Chapter 1

Non equilibrium theory

"Environmental entropy production is the price Nature has to pay for keeping a subsystem away from thermal equilibrium" [2]

1.1 Continuous Markov chains

Discrete time Markov chains evolve a probability distribution by one unit of time

$$\pi_i^{t+1} = \sum_j \pi_j^{t+1} P_{ji}$$

If we want to evolve the distribution of a time longer than one we can just use iteratively the relation above, obtaining:

$$\pi^{t+n} = \pi^t P^n$$

When we consider a continuous-time phenomenon we have to introduce a stochastic matrix $P(t)$ which depends on the continuous parameter t. We can ask ourself now how does the P matrix evolves with time. In order to do this we will study its derivative:

$$\frac{d}{dt}P(t) = \lim_{h\to 0} \frac{P(t+h) - P(t)}{h} \tag{1.1}$$

Using the Chapman Kolmogorov equation we have that $P(t+h) = P(t)P(h)$. Collecting $P(t)$:

$$\frac{d}{dt}P(t) = P(t) \lim_{h\to 0} \frac{P(h) - I}{h} = PQ \tag{1.2}$$

The Q matrix is called the "infinitesimal generator" of the Markov chain, and it would be largely used in this thesis. In particular it coincides with the rate matrix of a process described by a Master equation.
A very property of this matrix is that it gives also the time derivative of the

distribution vector π.

$$\frac{d}{dt}\pi(t) = \frac{d}{dt}(\pi(0)P(t)) = \pi(0)\frac{d}{dt}P(t) = \pi(0)P(t)Q$$

$$\frac{d}{dt}\pi(t) = \pi(t)Q \tag{1.3}$$

Another important property of matrix Q is that we can find matrix P from Q. Starting from equation (1.2)

$$\frac{\dot{P}}{P} = Q \tag{1.4}$$

which is easily solved, and we finally find the relation:

$$P(t) = e^{Qt} \tag{1.5}$$

Since in this thesis we would only be working with steady states, the $P(t)$ matrix will not be needed at all. The Q matrix containing the reaction rates will be largely used to find steady states (1.3). However it's important to understand the role of this matrix in the Markov chain framework.

1.1.1 Holding time

In discrete time markov chains random walkers change their location every unit of time (they can remain in their position if a selfloop is present.
On the contrary, in continuous time Markov chains, we have that the "holding time" is a random variable itself. Its distribution in time is exponential (exponential distributions are memoryless).
Considering a node i of a network, and calling k_{ij} the rates going out of node i, we have that the holding time distribution is:

$$P(t) = (\sum_j k_{ij}) exp\left(-(\sum_j k_{ij})t\right) \tag{1.6}$$

1.2 Master Equation

The master equation is a very simple yet extremely powerful equation that describes the time evolution of an enormous variety of different systems. Among the letter there are chemical reactions.
The time evolution equation of the concentration of a chemical species is expressed though the chemical rates.

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[X_c\right] = \sum_{c'}\left(\left[X_{c'}\right]w_{c'c} - \left[X_c\right]w_{cc'}\right) \tag{1.7}$$

This equation can easily be mapped in the master equation dividing the equation by the sum of all concentrations, thus getting a probability $P_c = \frac{P_c}{\sum_{c'}[X_{c'}]}$

$$\frac{\mathrm{d}}{\mathrm{d}t}P_c = \sum_{c'} \left(P_{c'} w_{c'c} - P_c w_{cc'}\right) \tag{1.8}$$

This is clearly equation (1.3). The generator Q of the Markov chain is a function of the rates. In physics and biophysics is very common to study steady states. For systems described by (1.8) they are found solving the (1.8) equation with a vanishing left hand side.

$$\sum_{c'} \left(P_{c'} w_{c'c} - P_c w_{cc'}\right) = 0 \tag{1.9}$$

1.2.1 Detailed Balance

The principle of Detailed Balance states that a system where elementary processes occur is in equilibrium when every of these processes is completely balanced by its reverse.
This principle gives a very clear definition of equilibrium and it is easily understood in terms of a Master Equation.
Considering (1.9) we have that Detailed Balance holds when all the terms in the sum are equal to zero:

$$P_{c'} w_{c'c} - P_c w_{cc'} = 0 \ \forall (c, c') \tag{1.10}$$

As (1.10) shows the detailed balance condition implies that there is no net current around all branches of the reaction graph.

All steady states that do not fulfill (1.8) are called Non Equilibrium Steady States (NESS) and they are for particular interest in Biophysics, given the intrinsic out of equilibrium nature of life.

1.2.2 Cycles

It is worth to notice that if Detailed Balance is respected and a cycle is present in the graph, the rates can not be completely independent. Using (1.10) recursively for the cycle edges, probabilities cancel out from the equation, leaving a relation that contains rates only:

$$\frac{\prod_i w_{i,i+1}}{\prod_i w_{i+1,i}} = 1 \tag{1.11}$$

Where the product is performed only on the nodes that belong to the cycle.
It's very important to clarify that this property does not assure that the system

is in equilibrium, but it says that the system has an equilibrium state to relax to. The relaxation toward equilibrium is an aspect of out of equilibrium physics which is not treated in this thesis.
Instead we will be dealing with "Non Equilibrium Steady States" (NESS) which are steady states that do not fulfill (1.11).

1.3 Useful definitions

1.3.1 Extent of a reaction

Lets consider a simple reaction, at constant temperature:

$$\mathrm{A} \rightleftharpoons \mathrm{B} \tag{1.12}$$

The following quantity is defined as the "extent of the reaction", and it measure the intensity of the reaction.

$$d\xi = dN_B = -dN_A \tag{1.13}$$

This quantity is often defined using particle concentration instead of particle numbers, but the meaning is the same.
The time evolution of this quantity is easily expressed via transition rates:

$$\dot{\xi} = N_A w_{AB} - N_B w_{BA} \tag{1.14}$$

ξ clearly vanish in equilibrium conditions.
The relation of the extent of the reaction with the Gibbs free energy is:

$$dG = (\mu_B - \mu_A)\, d\xi \tag{1.15}$$

which leads us to a new definition.

1.3.2 Chemical potential and chemical affinity

The chemical potential is defined as the change in Gibbs free energy as the particle number (or concentration) changes. Just as Gibbs free energy, it has both an energy and entropic term:

$$\mu_c = \frac{\partial F}{\partial N_c} = \mu_c^0 + k_B T \ln N_c \tag{1.16}$$

It is very important to notice that it is an increasing function of the specie concentration. For example a diffusion process can be explained in terms of chemical potentials. Particles migrate from high concentration areas to lower concentration

areas, thus releasing energy, and in turn making diffusion a spontaneous process (Fick's first law).
The difference between the chemical potentials belonging to dissimilar species is defined as the chemical affinity:

$$A_{cc'} = \mu_{c'} - \mu_c \tag{1.17}$$

$$A_{cc'} = \mu^0_{c'} - \mu^0_c + k_B T \ln \frac{N_{c'}}{N_c} = \Delta\mu^0_{cc'} + k_B T \ln \frac{N_{c'}}{N_c} \tag{1.18}$$

The chemical affinity is the change in chemical potential if we were to transform one particle of specie c to one of specie c'.

Setting the lhs of (1.18) to zero and solving for the concentration (number) ratio we get:

$$K = \frac{N'_c}{N_c} = e^{-\frac{\mu^0_{c'} - \mu^0_c}{kT}} \tag{1.19}$$

where K is called equilibrium constant of the reaction.

1.3.3 Smoluchowsky equation of chemical rates

Smoluchowski equation is a very simple, yet meaningful way to look at chemical rates.

$$K = \frac{k}{V} = \frac{4\pi \left(D_1 + D_2\right)\left(R_1 + R_2\right)}{V} = \frac{4\pi \hat{D} \sigma}{V} \tag{1.20}$$

This equation describes the rate at which two spherical particles, with radius R and diffusion constant D meet in a volume V. This is clearly linked to chemical reaction since they occur when to compatible molecules meet because of the random diffusion in the thermal bath.

1.4 Entropy production

In order to deeply understand Biolocical systems it's indispensable to study their relationship with the environment [3].

How can a biological system keep itself out of equilibrium, maintaining order, apparently violating the second principle of Thermodynamics?
Laws of physics tell us that all isolated systems are either in equilibrium or are tending to it, thermalizing. Biological systems cannot be isolated (they always depend on the environment for food, water and many other chemical compounds) and have to be considered subsystems, constantly interacting with the environment. The concept of Entropy production naturally arises as a way to reconcile NESS with

the second law of thermodynamics. A system can manage to keep itself ordered (for a finite time), in a low entropy state, but it will increase the environment entropy.

1.4.1 Operational fluxes

[4] In order to help us understand the concept of entropy production we will consider the example of particles going in and out of a biological subsystem. Introducing an extensive parameter X_c (could be the concentration of the particle). Calling X_c^s the parameter for the subsystem and X_c^e the one for the environment we can write:

$$X_c = X_c^s + X_c^e \tag{1.21}$$

This could be the case of the transport of a specific molecule in or out of a cell. Taking the derivative of Gibbs potential on variable X, and assuming that both internal energy and volume do not depend on this parameter, we get:

$$F_c = \frac{dG}{dX_c} = \frac{dS}{dX_c} = \frac{d(S^s + S^e)}{dX_c} = F_c + F_c' \tag{1.22}$$

F_c is called the "thermodynamic force" or affinity. Most of the times it does not have the unit of a force, but the analogy with a force is in the fact it is that it is the derivative of a potential on a certain variable ($F = -\partial_x U$). Clearly the system is at equilibrium only when there is no net force acting on it.
Introducing the flux J_c:

$$J_c = \frac{dX_c}{dt} \tag{1.23}$$

It's important to note that this fluxes can exist even if the subsystem is in a steady state, indeed:

$$J_c = \frac{dX_c}{dt} = \frac{d(X_c^s + X_c^e)}{dt} = \frac{d(X_c^e)}{dt} \tag{1.24}$$

Where J_c is called the "operational flux" of particle specie "c" going in the system.

$$\frac{dS}{dt} = \sum_c \frac{dS}{dX_c}\frac{dX_c}{dt} = \sum_c F_c^{eq} J_c \tag{1.25}$$

This relation can be interpreted as follows: J_c is the number of particles that going in the subsystem. F_c is the free energy increase/decrease associated to every particle that crosses the system's barrier. More specifically F_c is the difference in the chemical potentials associated to the inside and outside particles.

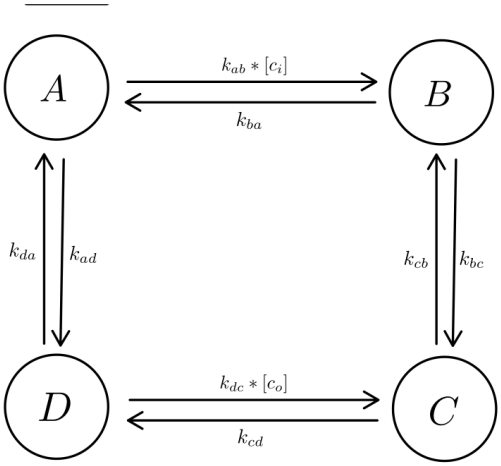
$$F_c = \mu_c - \mu_c' \tag{1.26}$$

Operational fluxes, being the time derivative of a concentration, can be experimentally measured. We will see that this is not the case for the cycle-fluxes.

1.4.2 Cycle-fluxes

In this section we will observe the same kind of phenomena discussed in the previous section but from the perspective of cycle dynamics.
We will consider a simple system that transports a specific molecule inside and outside of the molecule.
This cycle brings particles outside if considered in the counter-clockwise direction,

Figure 1.1: Reaction graph for a generic process



and brings one particle inside the system if for every cycle completed in the clockwise direction.
From (1.11) we now that if the product of the clockwise rates is equal to the counterclockwise rates, the system can achieve equilibrium. We will cal this ratio Ω and will understand its meaning.

$$\Omega = \frac{w_{ab}k_{bc}k_{cd}k_{da}}{k_{ba}k_{cb}w_{dc}k_{ad}} \tag{1.27}$$

Recalling that both w_{ab} and w_{dc} are rates associated to second order reaction (the flux depends on 2 concentrations). Therefore they also depend on the (respectively) internal/external concentration of the molecule. Therefor Ω is a function of the concentrations:

$$\Omega = \frac{(c_i k_{ab})k_{bc}k_{cd}k_{da}}{k_{ba}k_{cb}(c_o k_{dc})k_{ad}} \tag{1.28}$$

at equilibrium we have $\Omega = 1$, therefore:

$$(\frac{c_i^{eq}}{c_o^{eq}})^{-1} = \frac{k_{ab}k_{bc}k_{cd}k_{da}}{k_{ba}k_{cb}k_{dc}k_{ad}} \tag{1.29}$$

Substituting back in (1.28) we get that Ω is just the ratio between the ratios of the concentrations:

$$\Omega = \frac{\left(\frac{c_i}{c_o}\right)}{\left(\frac{c_i}{c_o}\right)_{eq}} \tag{1.30}$$

Recalling the definition of the chemical potential (1.16) and the one of the affinity we have:

$$\frac{c_i}{c_o} = \exp\left(\frac{F - \Delta\mu^0}{k_B T}\right) \tag{1.31}$$

Using the same scheme for the ratio at equilibrium (F=0) and substituting in (1.30) we finally link the Ω to the thermodynamic force:

$$\Omega = \exp\left(\frac{F}{k_B T}\right) \tag{1.32}$$

$$F = k_B T \log \Omega = k_B T \log\left(\frac{\left(\frac{c_i}{c_o}\right)}{\left(\frac{c_i}{c_o}\right)_{eq}}\right) \tag{1.33}$$

This simple example shows that, if second order rates are present, we can express Ω as a function of these concentrations, and the ones at equilibrium.

Chapter 2

Clamping and Sliding model for the ParBS system

ParABS systems ensure chromosome separation for most bacterial species. This chapter will focus on one component of this tripartite system: the ParB protein, and in particular on the conformational changes it can go through. This dimer can transition between 2 shape conformation (open and closed), and it can bind to the both to the cytidine triphosphate (CTP) and to the cytidine diphosphate (CDP) molecules. The role of these nucleotides is crucial for the thermodynamical aspects of this system: CTP hydrolysis provides the energy necessary for keeping the system out of equilibrium.

Figure 2.1: Open state of the dimer [5]

2.1 The CTP solution

In order to understand ParB conformational changes it's important to understand its relationship and interaction with the solution it is found in. In particular we

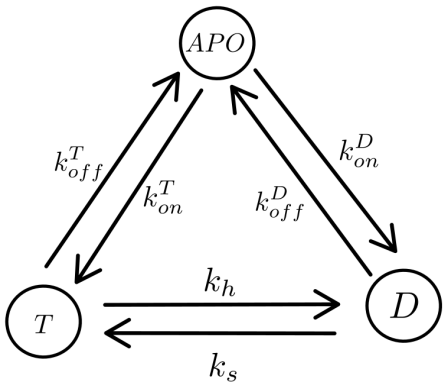
Figure 2.2: Closed state of the dimer [5]

are interested on the concentration of the species of the 2 nucleotides, CDP and CTP. These two nucleotides can transform into each other through the processes of hydrolysis and synthesis. The rate of CTP synthesis k_s is much lower than the rate of hydrolysis k_h, indeed the ratio of the equilibrium concentrations is:

$$\frac{[CDP]_{eq}}{[CTP]_{eq}} \approx 10^6$$

However this is often not the case of the solution in the cell, where CTP is very abundant, way more than CDP.
The ParB dimer can find itself bound to one of the two nucleotides (T and D states in the next graph) or bound to no nucleotide, the APO state. We will show that this state can be ignored since its thermodynamically unfavourable.

Figure 2.3: Reaction graph for off-DNA ParB's states



Let's now consider the reaction graph fig.2.3. It is the reaction chain that a ParB dimer undergoes when found in a solution of CTP and CDP.

Considering the ratio of clockwise and counter clockwise rates we get:

$$\Omega = \frac{k_h k_{off}^D k_{on}^T [CTP]}{k_s k_{off}^T k_{on}^D [CDP]}$$

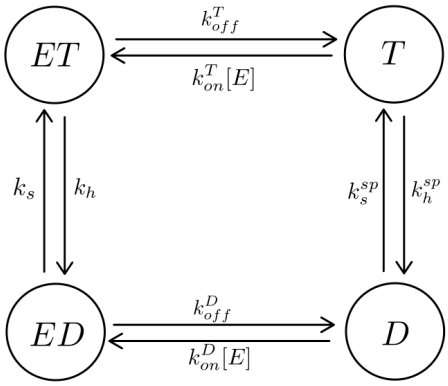
Introducing the dissociation constant: $k_d^i = \frac{k_{off}^i}{k_{on}^i}$ we get:

$$\Omega = \frac{k_h k_d^D [CTP]}{k_s k_d^T [CDP]} \tag{2.1}$$

This ratio is known to be equal to one (the thermodynamic force of the cycle is zero) if the rates permit an equilibrium steady-state. From (2.1) it's evident that equilibrium is only possible at a specific value of the $[CTP]/[CDP]$ ratio.

Now we will consider the reactions that a CTP or CDP molecule undergoes in the presence of a concentration $[E]$ of ParB molecule, treated in this section as an enzyme. k_s and k_h are the same rates of fig.2.3 while k_s^{sp} and k_h^{sp} are the spontaneous hydrolysis and synthesis rates of the solution (nucleotides free, not bound the the protein). It's reasonable to assume that this cycle is able to achieve

Figure 2.4: Reaction graph for a CTP/CDP molecule



equilibrium. For this reason we set the thermodynamic force to zero, obtaining:

$$\frac{k_s k_{off}^T k_h^{sp} k_{on}^D [E]}{k_h k_{on}^T [E] k_s^{sp} k_{off}^D} = \frac{k_s k_d^T k_h^{sp}}{k_h k_d^D k_s^{sp}} = 1 \tag{2.2}$$

It's important to notice that the equilibrium condition does not depend on the enzime concentration. Manipulating the former expression we get:

$$\frac{k_s^{sp}}{k_h^{sp}} = \frac{k_s k_d^T}{k_h k_d^D} \tag{2.3}$$

But the lhs can be expressed in terms of equilibrium concentrations. Imposing detailed balance over the hydrolisis/synthesis reaction we have:

$$\frac{k_s^{sp}}{k_h^{sp}} = \left(\frac{[CTP]}{[CDP]}\right)_{eq} \tag{2.4}$$

Combining (**??**) with (2.3) we obtain:

$$\frac{k_s k_d^T}{k_h k_d^D} = \left(\frac{[CTP]}{[CDP]}\right)_{eq} \tag{2.5}$$

finally inserting (2.5) in (2.1):

$$\Omega = \frac{\frac{[CTP]}{[CDP]}}{\left(\frac{[CTP]}{[CDP]}\right)_{eq}} \tag{2.6}$$

It's now evident that this cycle needs the equilibrium concentration ratio of CTP/CDP in order to be at equilibrium itself.

2.2 The exchange rates

Recalling fig.2.3, we will now describe the system by means of a master equation. In this section we will consider the time evolution of the enzyme concentration by means of a Master equation:

$$\begin{cases} \dot{[a]} = & +k_{off}^T[ET] - k_{on}^T[T][a] + k_{off}^D[ED] - k_{on}^D[D][a] \\ \dot{[ET]} = & -k_h[ET] + k_s[ED] - k_{off}^T[ET] + k_{on}^T[T][a] \\ \dot{[ED]} = & +k_h[ET] - k_s[ED] - k_{off}^D[ED] + k_{on}^D[D][a] \end{cases} \tag{2.7}$$

Previous publications [5] show that nucleotide binding and unbinding rates are much faster than the Hydrolysis/Synthesis one. Therefore we can consider the APO state (whose master equation contains the fast rates only) constant in time. The next calculation will show how to permorm a coarse-graining of the system where the short-lived APO state is neglected.
Exploiting the stationarity condition we are able to find $[a]$ as a function of the other 2 concentrations:

$$[a] = \frac{k_{off}^T[ET] + k_{off}^D[ED]}{k_{on}^T[T] + k_{on}^D[D]} \tag{2.8}$$

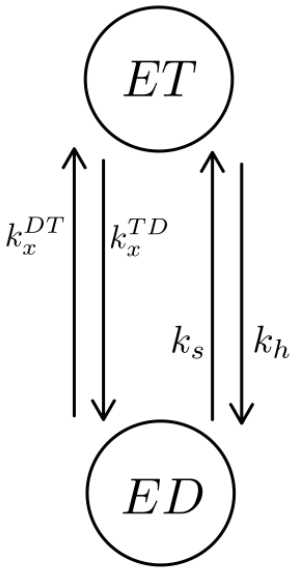
substituting this relation in (2.7) and reorganizing terms we obtain 2 equations where a new interaction (between ED and ET) has appeared: the exchange interaction.

$$k_x^{D\to T} = k_{off}^D \frac{k_{on}^T[CTP]}{k_{on}^T[CTP] + k_{on}^D[CDP]} = \frac{k_{off}^D k_{on}^T \frac{[CTP]}{[CDP]}}{k_{on}^T \frac{[CTP]}{[CDP]} + k_{on}^D} \tag{2.9}$$

$$k_x^{T\to D} = k_{off}^T \frac{k_{on}^D[CDP]}{k_{on}^T[CTP] + k_{on}^D[CDP]} = \frac{k_{off}^T k_{on}^D}{k_{on}^T \frac{[CTP]}{[CDP]} + k_{on}^D} \tag{2.10}$$

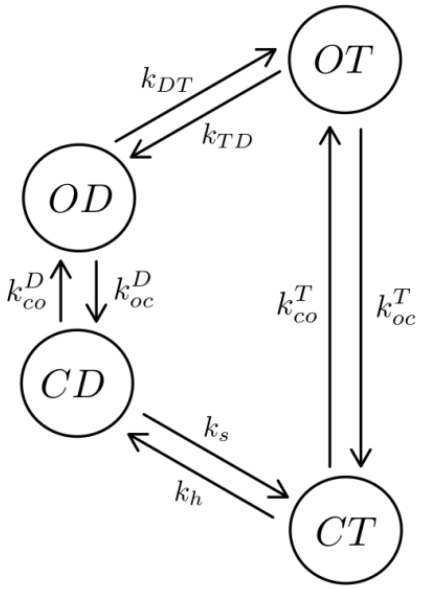
This new interaction originates from the the removal of the APO state. The system in fig.2.3 can now be seen as a 2 state systems with additional edges:sythesys It's

Figure 2.5: Coarse grained ParB state transition graph



now important to notice that the exchange reactions (in both directions) are still much faster than the hydrolysis/synthesis.
When the dimer is open fig.(2.2) the nucleotide is free to transform both with exchange and hydrolysis/synthesis, so from now one we will consider only the exchange process since it's the dominant one.
On the contrast it has been observed that when the dimer closes it "traps" the nucleotide fig.(2.2) therefore making the exchange not possible. In the graph below

Figure 2.6: The 4 possible states of the dimers and the relative rates



2.3 The thermodynamic force

Once again we consider the ratio between the two directions of a chemical cycle. We now consider the cycle of fig.**??**

$$\Omega_4 = \frac{k_h k_{co}^D k_x^{D\to T} k_{oc}^T}{k_s k_{oc}^D k_x^{T\to D} k_{co}^T} \tag{2.11}$$

Substituting the expression of the exchange rates (2.9) and (2.10) and recalling the definition of the dissociation constants:

$$\frac{k_x^{D\to T}}{k_x^{T\to D}} = \frac{k_d^D}{k_d^T}\alpha \tag{2.12}$$

where $\alpha = \frac{[CTP]}{[CDP]}$.
Using relation (2.5):

$$\frac{k_x^{D\to T}}{k_x^{T\to D}} = \frac{\alpha k_s}{\alpha_{eq} k_h}$$

We finally obtain an expression that links the thermodynamical force of this cycle to the nucleotide's concentrations:

$$\Omega_4 = \frac{\alpha k_{co}^D k_{oc}^T}{\alpha_{eq} k_{oc}^D k_{co}^T} \tag{2.13}$$

Chapter 3

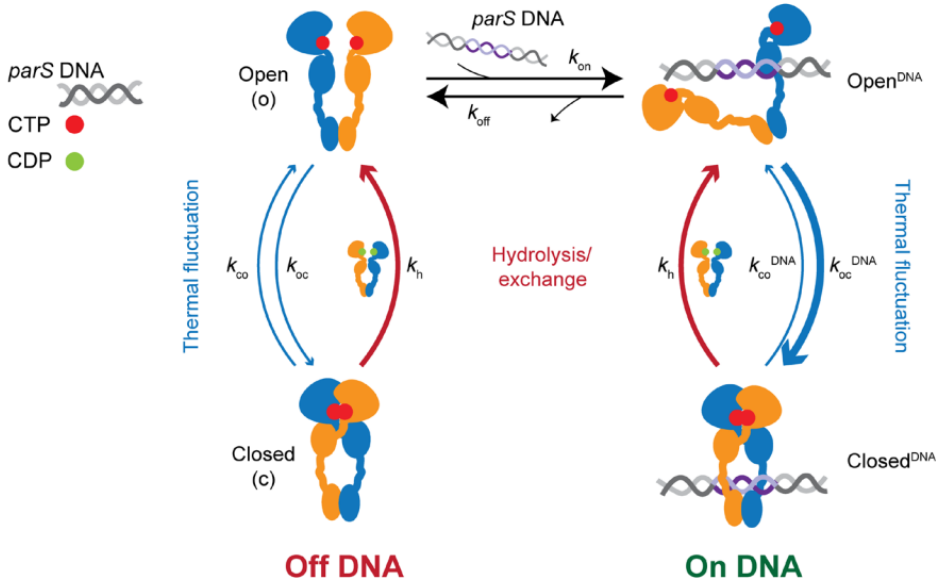
From a 5-state system to a 3-state one

The 5-state model is the one that is thought to be the closest to the real biological model. Such model is however impossible to solve analytically even in a continuous limit. A continuous 2 state model (empty/full) has already been formulated [1] in a phenomenological fashion. A reasonable trade-off between the the desire of an analytical solution and a less coarse grained description of the system is achieved with a 3 state model (empty, full-open, full-closed). The letter would be solvable (in the continuous limit) and it would contain opening and closure rates, indispensable for an accurate description of the ParS-ParB interaction. In this chapter a dynamical coarse-graining procedure has been developed, which, remarkably, is able to preserve the thermodynamic force of the cycle.

3.1 Justification for the coarse-graining

The idea of this coarse-graining procedure is to get rid of 2 out of the 4 conformations: the open-CDP and the closed-CDP states. This is done because it has been observed that both these configurations are thermodynamically unstable and thus are short-lived [5]. The OD state is considered transient since the CTP concentration is orders of magnitude higher than the CDP (therefore we have $k_{DT} >> k_{TD}$). The reason why the CD state is short-lived is on the contrast due to the fact that the the ParB dimer is prone to open itself when it is bound to CDP. As depicted in fig.(3.1) the OD and CD state will be considered as "transition-states" or "reactive intermediates". The coarse-grained model would than be made just out of the two states bound to CTP and the effective rates between them will be the inverse of the mean passage times that also include the presence of the CDP-bound states.

Figure 3.1: A reaction graph showing the main reaction both off-DNA and on-DNA [5]



3.2 Mean passage time approach

In order to perform the sought coarse-graining we will use the concept of mean passage times. For a discrete time Markov chain it reads:

$$m_{ij} = p_{ij} + \sum_{k \neq j} p_{ik}(1 + m_{kj}) \tag{3.1}$$

When time is discrete the random walker waits exactly 1 unit of time between every jump. This relation is clearly a recursive relation that explores every possible path, even infinite ones, to state j.
Switching to continuous-time Markov chains we need to introduce time dependent probability distributions. Recalling (1.6) we know how the functional form of single node holding-time distributions, however combining them and generalizing (3.1) is not trivial.

$$P_{ij}(t) = p_{ij}P_i(t) + \int_0^t p_{ik}P_k(t')p_{kj}P_j(t'')\delta(t - t' - t'')\,dt'dt''... \tag{3.2}$$

Where p_{ij} is the ratio between rate k_{ij} and the sum of all other rates going out of node i. Every term of the infinite sum is a convolution. Using the property of the Laplace transform that the transform of the convolution of 2 or more functions is

the product of the transforms of these functions:

$$\tilde{P_{ij}}(s) = p_{ij}\tilde{P}_i(s) + p_{ik}\tilde{P}_k(s)p_{kj}\tilde{P}_j(s)... \tag{3.3}$$

The sum is infinite, and if the chain doesn't have any sink the probability of reaching node j from node i in an infinite time must be equal to 1. Recalling the definition of the Laplace transform:

$$\tilde{P}_i(s) = \int_0^\infty P_i(t)e^{-st}\,dt \tag{3.4}$$

We can find the functional form of this Laplace transforms:

$$\tilde{P}_i(s) = \lambda_i \int_0^\infty e^{-\lambda_i t}e^{-st}\,dt = \frac{\lambda_i}{\lambda_i + s} \tag{3.5}$$

and evaluating it in $s = 0$ we have:

$$\tilde{P}_i(s=0) = \int_0^\infty P_i(t)e^0\,dt = 1 \tag{3.6}$$

Now that we know all the ingredients in (3.3) we know the pdf of the i-j transition. Finding the mean passage time it's now only a matter of taking an expectation value:

$$\tau_{ij} = \int_0^\infty tP_{ij}(t)\,dt = -\partial_s\tilde{P}_{ij}(s=0) \tag{3.7}$$

Once obtained the mean passage time, we can define the coarse-grained rate as the inverse of the mean passage time of the i-j transition.

$$k_{ij} = \frac{1}{\tau_{ij}} \tag{3.8}$$

3.3 Recursive method

Now we can proceed to calculate all the terms of the sum (3.3).

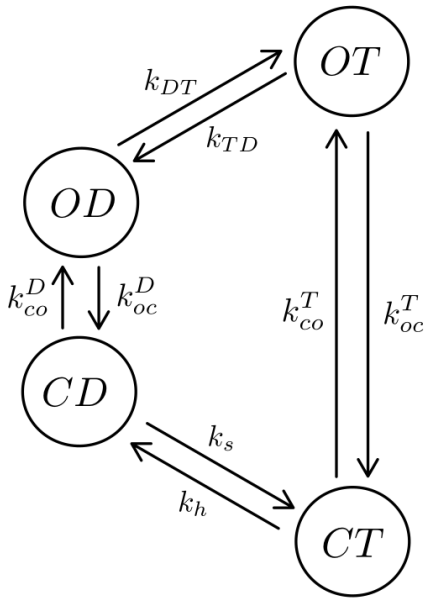
First we want to separate the infinite sum (3.3) in two subsets, for example for $P_{co}(t)$:

$$\tilde{P}_{co}(s) = \tilde{P}^D_{co}(t) + \tilde{P}^T_{co}(t) \tag{3.9}$$

Where the first term is the sum of all the paths that do not use edge k^T_{co} while the second is the sum of all terms that do. Let's focus on $P^T_{co}(t)$.
The first contribution to the sum is the shortest path to state CT:

$$P^T_{co}(t) = p^T_{co}P_{CT}(t) + ... \tag{3.10}$$

Figure 3.2: The five states model diagram (with no empty state)



where $p_{co}^T = \frac{k_{co}^T}{k_{co}^T + k_h} = \frac{k_{co}^T}{k_{CT}}$

The second term corresponds to the path that goes to state CD, back to CT and finally to OT. It's important to notice that the cycle CD-CT can be repeated infinite times. Another possible cycle is the jumping back and forth from state CD to OD. We are almost ready to write the recursive relation that solves our sum, defining some functions to make notation lighter:

$$z = p_h p_s \tilde{P}_{CT}(s) \tilde{P}_{CD}(s)$$

$$q = p_{oc}^D p_{co}^D \tilde{P}_{CD}(s) \tilde{P}_{OD}(s)$$

$$\tilde{P}_{co}^T(s) = p_{co}^T \tilde{P}_{CT}(s) + z \left(\sum_{n=0}^{\infty} q^n \right) \tilde{P}_{co}^T(s) \tag{3.11}$$

In this equation the memory-less property of Markov processes was used: once the path is back in the initial state, no matter how long it already has been, the probability $\tilde{P}_{co}^T(s)$ is unchanged, and therefore we can multiply by it. It's important to notice that the second term in the sum is actually a collection of infinite terms: each one has a different integer power of q . This can be understood in the following way: if a cycle of type z is performed, in the middle of it infinite cycles of type q

can happen. Solving for $\tilde{P}^T_{co}(s)$we get:

$$\tilde{P}^T_{co}(s) = p^T_{co}\tilde{P}_{CT}(s)\frac{1}{1-\frac{z}{1-q}} \tag{3.12}$$

$\tilde{P}^D_{co}(s)$ has a slightly different expression. In this case the shortest path is made out of 3 edges: k_h k^D_{co} k_{DT}. The recursive relation reads:

$$\tilde{P}^D_{co}(s) = +z\left(\sum_{n=0}^{\infty} q^n\right)\tilde{P}^D_{co} + p_h\tilde{P}_{CT}p^D_{co}\tilde{P}_{CD}p_{DT}\tilde{P}_{OD}\left(\sum_{n=0}^{\infty} q^n\right) \tag{3.13}$$

Solving it, we obtain:

$$\tilde{P}^D_{co}(s) = \frac{1}{1-q-z}p_h\tilde{P}_{CT}p^D_{co}\tilde{P}_{CD}p_{DT}\tilde{P}_{OD}$$

The recursive relations for the $o \to c$ rates are very similar and follow the same scheme. As explained in the previous section we now find the mean passage time as a derivative of the Laplace transform of the the pdf.

$$\tau_{co} = -\partial_s\tilde{P}_{co}(s)|_{s=0} = -\partial_s(\tilde{P}^D_{co}(s) + \tilde{P}^T_{co}(s))|_{s=0} \tag{3.14}$$

$$\tilde{k}^D_{co} = \tau^{-1}_{co}\tilde{P}^D_{co}(0) \tag{3.15}$$

$$\tilde{k}^T_{co} = \tau^{-1}_{co}\tilde{P}^T_{co}(0) \tag{3.16}$$

For the $c \to o$ rates we finally get:

$$\tilde{k}^T_{co} = \frac{k_{\mathrm{co},T}\left(k_{\mathrm{DT}}\left(k_{\mathrm{co},D}+k_s\right)+k_sk_{\mathrm{oc},D}\right)}{k_{\mathrm{DT}}\left(k_{\mathrm{co},D}+k_h+k_s\right)+k_h\left(k_{\mathrm{co},D}+k_{\mathrm{oc},D}\right)+k_sk_{\mathrm{oc},D}} \tag{3.17}$$

$$\tilde{k}^D_{co} = \frac{k_{\mathrm{DT}}k_hk_{\mathrm{co},D}}{k_{\mathrm{DT}}\left(k_{\mathrm{co},D}+k_h+k_s\right)+k_h\left(k_{\mathrm{co},D}+k_{\mathrm{oc},D}\right)+k_sk_{\mathrm{oc},D}} \tag{3.18}$$

Instead for $o \to c$ we have:

$$\tilde{k}^T_{oc} = \frac{k_{\mathrm{oc},T}\left(k_{\mathrm{DT}}\left(k_{\mathrm{co},D}+k_s\right)+k_sk_{\mathrm{oc},D}\right)}{k_{\mathrm{DT}}\left(k_{\mathrm{co},D}+k_s\right)+k_{\mathrm{TD}}\left(k_{\mathrm{co},D}+k_{\mathrm{oc},D}\right)+k_s\left(k_{\mathrm{oc},D}+k_{\mathrm{TD}}\right)} \tag{3.19}$$

$$\tilde{k}^D_{oc} = \frac{k_sk_{\mathrm{TD}}k_{\mathrm{oc},D}}{k_{\mathrm{DT}}\left(k_{\mathrm{co},D}+k_s\right)+k_{\mathrm{TD}}\left(k_{\mathrm{co},D}+k_{\mathrm{oc},D}\right)+k_s\left(k_{\mathrm{oc},D}+k_{\mathrm{TD}}\right)} \tag{3.20}$$

3.4 The ParS rates

In experiments [5] it is observed that the ParS gene in DNA manages to catalyze the opening-closure reaction. The presence of this catalysis site is essential for the diffusion of ParB dimers over DNA and in turn for chromosome separation in bacteria. Fig. 3.1 is an explicative figure showing the main reactions at play.
For this reason in this model we will consider that these rates are scaled by a factor γ_D and γ_T respectively for the D and T states. All the other reaction rates will remain the same.
The coarse-graining procedure is exactly the same of the one for the non specific rates.

$$\tilde{Sk}_{co}^{T} = \frac{\gamma_T k_{\mathrm{co},T}\left(k_{\mathrm{DT}}\left(\gamma_D k_{\mathrm{co},D} + k_s\right) + \gamma_D k_s k_{\mathrm{oc},D}\right)}{k_{\mathrm{DT}}\left(\gamma_D k_{\mathrm{co},D} + k_h + k_s\right) + \gamma_D\left(k_h\left(k_{\mathrm{co},D} + k_{\mathrm{oc},D}\right) + k_s k_{\mathrm{oc},D}\right)} \tag{3.21}$$

$$\tilde{Sk}_{co}^{D} = \frac{\gamma_D k_{\mathrm{DT}} k_h k_{\mathrm{co},D}}{k_{\mathrm{DT}}\left(\gamma_D k_{\mathrm{co},D} + k_h + k_s\right) + \gamma_D\left(k_h\left(k_{\mathrm{co},D} + k_{\mathrm{oc},D}\right) + k_s k_{\mathrm{oc},D}\right)} \tag{3.22}$$

$$\tilde{Sk}_{oc}^{T} = \frac{\gamma_T k_{\mathrm{oc},T}\left(k_{\mathrm{DT}}\left(\gamma_D k_{\mathrm{co},D} + k_s\right) + \gamma_D k_s k_{\mathrm{oc},D}\right)}{k_{\mathrm{DT}}\left(\gamma_D k_{\mathrm{co},D} + k_s\right) + \gamma_D k_{\mathrm{TD}}\left(k_{\mathrm{co},D} + k_{\mathrm{oc},D}\right) + k_s\left(\gamma_D k_{\mathrm{oc},D} + k_{\mathrm{TD}}\right)} \tag{3.23}$$

$$\tilde{Sk}_{oc}^{D} = \frac{\gamma_D k_s k_{\mathrm{TD}} k_{\mathrm{oc},D}}{k_{\mathrm{DT}}\left(\gamma_D k_{\mathrm{co},D} + k_s\right) + \gamma_D k_{\mathrm{TD}}\left(k_{\mathrm{co},D} + k_{\mathrm{oc},D}\right) + k_s\left(\gamma_D k_{\mathrm{oc},D} + k_{\mathrm{TD}}\right)} \tag{3.24}$$

Clearly the structure of these expression is the same of the NS rates, with γ factors showing up every time that an opening-closure rate shows up.

3.4.1 Thermodynamic force after the coarse-graining

Plugging in the previous expressions it's trivial to check that the thermodynamical force of cycle (fig.3.2) is the same of the coarse grained system.

$$\frac{k_s k_{\mathrm{TD}} k_{\mathrm{co},T} k_{\mathrm{oc},D}}{k_{\mathrm{DT}} k_h k_{\mathrm{co},D} k_{\mathrm{oc},T}} = \frac{\tilde{k}_{co}^{T}\tilde{k}_{oc}^{D}}{\tilde{k}_{oc}^{T}\tilde{k}_{co}^{D}} \tag{3.25}$$

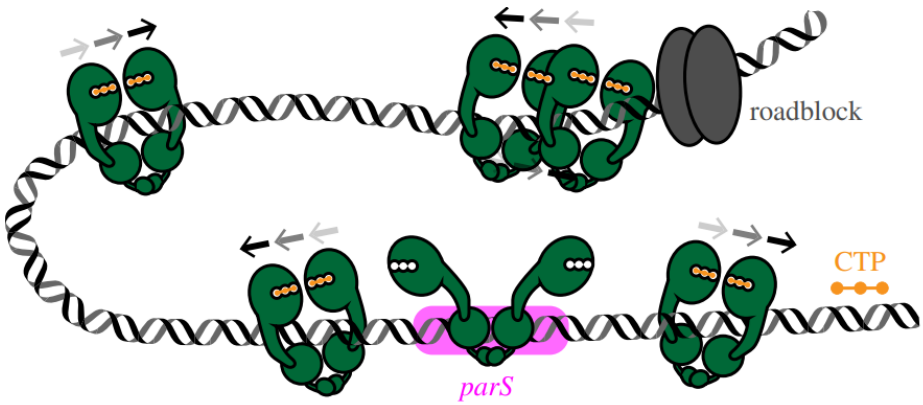
This is very important from an out-of-equilibrium perspective: the system, even if coarse grained, dissipates the same amount of energy every time it completes a cycle.

Chapter 4

The discrete model and the continuous limit

While in chapter 2 the conformational changes of the ParB dimer have been explained, in this chapter a model will be built in order to describe the interaction and spatial distribution of these dimers on DNA. Many qualitative models have been proposed in [6] to explain this spatial distribution. This thesis focuses on the model known as "Clamping and Sliding" which describes the tendency of the ParB dimers to attach to DNA (clamping) and than move laterally on it (sliding).

Figure 4.1: An image showing the clamp and slide model [6]



4.1 The discrete model

The model proposed in this thesis considers a long chain of sites (the DNA). Every site of the lattice can be in one of these 3 states: empty (e), occupied by an open ParB dimer (o) and occupied by a closed ParB dimer (c).

$$P_i^o + P_i^c + P_i^e = 1$$

In this section the model will be solved in the most general fashion. Later biology-based simplifications will be made in order to have a better understanding of the solution.
Dimers which are present in the solution can occupy an empty site with rate k_{on}^o or k_{on}^c. This attaching flux depends on the ParB concentration in the solution surrounding the DNA. The relation is the following:

$$J = k_{on}[ParB]P_i^e = w_{on}P_i^e \tag{4.1}$$

At the same way dimers can detach from DNA with rates k_{off}^o and k_{off}^c.
Both open and closed ParB dimers can diffuse on the chain. The hopping rate between sites is respectively q_o and q_c for the two states.
Finally dimers are also allowed to transition from the open to closed state with rates k_{oc} and k_{co}.

$$\begin{cases} \dot{P}_i^o = \frac{q_o}{2}(P_{i+1}^o + P_{i-1}^o) - q_o P_i^o + k_{co}P_i^c - k_{oc}P_i^o - k_{off}^o P_i^o + w_{on}^o P_i^e \\ \dot{P}_i^c = \frac{q_c}{2}(P_{i+1}^c + P_{i-1}^c) - q_c P_i^c - k_{co}P_i^c + k_{oc}P_i^o - k_{off}^c P_i^c + w_{on}^c P_i^e \\ \dot{P}_i^e = -\frac{q_o}{2}(P_{i+1}^o + P_{i-1}^o) + q_o P_i^o + k_{off}^o P_i^o - k_{on}^o P_i^e - \frac{q_c}{2}(P_{i+1}^c + P_{i-1}^c) + q_c P_i^c + k_{off}^c P_i^c - w_{on}^c P_i^c \end{cases} \tag{4.2}$$

Taking the time derivative of the normalization property of probability we have that $\dot{P}_i^e = 1 - \dot{P}_i^o - \dot{P}_i^o$ clearly meaning that the equation for the empty state it's a linear combination of the other two equations. Substituting $P^e = 1 - P^o - P^c$, rearranging terms and writing the two equations in a compact vector notation:

$$\dot{\vec{P}}_i = \frac{Q}{2}(\vec{P}_{i+1} + \vec{P}_{i-1} - 2\vec{P}_i) - W_{tot}(i)\vec{P}_i + W_{on}\vec{P}_i + \vec{c} \tag{4.3}$$

with:

$$Q = \begin{pmatrix} Q_o & 0 \\ 0 & Q_c \end{pmatrix} \tag{4.4}$$

and

$$-W_{tot} = \begin{pmatrix} -k_{off}^o - k_{oc} & +k_{co} \\ +k_{oc} & -k_{off}^c - k_{co} \end{pmatrix} \tag{4.5}$$

$$W_{on} = \begin{pmatrix} -w_{on}^o & -w_{on}^o \\ -w_{on}^c & -w_{on}^c \end{pmatrix} \tag{4.6}$$

$$\vec{c} = \begin{pmatrix} w_{on}^{o} \\ w_{on}^{c} \end{pmatrix} \tag{4.7}$$

4.1.1 The opening-and closure rates

Experimental data show a very different behaviour of ParB dimers around the ParS gene. The latter is positioned in the origin of the system. In order to emphasize this different behaviour of the system in the origin we introduce W, being the matrix of the rates that holds everywhere except in the origin (in the non-specific DNA)

$$W_{tot} = W + (W_{tot} - W) = W + W_0 \tag{4.8}$$

Noticing that W_0 is zero everywhere except in the origin (ParS site), therefore we can write:

$$W_{tot} = W + \delta_{i,0} W_0 \tag{4.9}$$

Finally we obtain:

$$\dot{\vec{P}}_i = \frac{Q}{2}(\vec{P_{i+1}} + \vec{P_{i-1}} - 2\vec{P}_i) - (W + \delta_{i,0} W_0)\vec{P}_i + W_{on}\vec{P}_i + \vec{c} \tag{4.10}$$

4.2 Continuous limit

The justification for the decision to take a continuous limit of the previously discussed discrete model is two-fold: the first one is to have an analytical solution whose parameter can be easily tuned and their relation studied. The second one is due to the desire to connect with [1] and propose a model which is a generalization of the one previously proposed in the literature. The former model is very coarse grained and it describes the distribution of dimers on DNA regardless their shape or nucleotide they are bound to (in experiments it is not possible to observe the shape of the dimer). The diffusion equation that governs this coarse grained picture has only few parameters:

$$\frac{\partial \rho(x,t)}{\partial t} = D\Delta\rho(x,t) + R\delta(x) - \rho(x,t)U \tag{4.11}$$

D is the diffusion constant, U is the unbinding constant, uniform on all DNA and finally R is a parameter that quantifies the strength of the source, therefore creating a spike around this source which is the ParS sequence of DNA.
This equation tries to fit experimental density distribution found in [7]
The model developed in this thesis aims to describe this system more faithfully. In order to do so we will consider a less coarse grained system where we are still able to distinguish between open and closed conformations of the dimer. This new degree of freedom permits of to characterize the effect of the ParS on the distribution of the ParB proteins (the creation of large clusters around the sequence) in a different way. This more general way to write the model will be explored in the next chapters.

In order to perform the continuous limit we will consider that each site has size a and on each of them there is a constant continuous distribution ρ_i such that:

$$P_i = a\rho_i \tag{4.12}$$

Substituting the former relation in the equation, and dividing by a

$$\dot{\vec{\rho}}_i = \frac{Qa^2}{2}\left(\frac{\vec{\rho_{i+1}} + \vec{\rho_{i-1}} - 2\vec{\rho_i}}{a^2}\right) - (W + \delta_{i,0}W_0)\vec{\rho_i} + W_{on}\vec{\rho_i} + \frac{\vec{c}}{a} \tag{4.13}$$

Setting the lattice site size a to zero we get the following equation:

$$\dot{\vec{\rho}} = D\vec{\rho} - (W + f_0(x)W_0)\vec{\rho} + w_{on} \tag{4.14}$$

In the next subsections the continuous limit will be fully explained.

4.2.1 Space derivative and diffusion constant

The first term is the diffusive part of the master equation. When a vanishes we get the continuous diffusion constant.

$$D_o = \lim_{a\to 0} \frac{Q_o a^2}{2}$$

$$D_c = \lim_{a\to 0} \frac{Q_c a^2}{2}$$

The parenthesis in the first term is the second spatial derivative of the density function.

4.2.2 The w_{on} rates

It is really important to focus on the last two terms of the equation. For the open state they read:

$$-w^o_{on}\rho^o_i - w^o_{on}\rho^c_i + \frac{w^o_{on}}{a} \tag{4.15}$$

It is crucial to analyze the dependence of w_{on} on a. This dependence can be understood by means of Smoluchowski equation for chemical rates. From (**??**) we understand that this rate depends on the sum of the reacting particle size. Performing the continuum limit and therefore setting a to zero means to set both the DNA site (particle 1) and the dimer (particle 2) size to zero. This means that $w_{on} \propto a$.

$$\lim_{a\to 0} -w^o_{on}\rho^o_i - w^o_{on}\rho^c_i + \frac{w^o_{on}}{a} = \tilde{w}^o_{on} \tag{4.16}$$

We have two consequences from this result: the first one is that $\tilde{w}^o_{on}$ is now a density (for the sake of short notation this density will be called w_{on}), the second one is that the entire W_{on} matrix vanishes in the continuum limit. An even even deeper consequence is that in the continuous limit an interaction is lost: in the discrete model, when a site has an high density (sum of open and closed almost equal to one) this would result in an overall lower binding-rate. This does not happen in the continuum limit: the w_{on} rate is the same everywhere regardless of the density.

4.3 The Kronecker delta term

The term multiplied by the Kronecker delta is incredibly important because it is the one linked to the interaction of the proteins with the ParS sequence. How this term is treated in the continuous limit is both delicate and crucial for this model. Two main approaches have been tried in this thesis.

4.3.1 Finite size ParS

A different approach is to consider ParS a finite region of DNA. This approach basically divide DNA in two parts, the centromere region (whose length is denoted with d), and the rest of DNA, often called Non Specific DNA. In these two regions the equations have the same structure but they have different opening/closure rates. This approach will be investigated in the next chapter.
This way to rewrite the model is believed to be more general since it should be possible to find the previous approach as a limit to $d \to 0$. So far computing this limit has been attempted but it proved challenging.

4.3.2 Infinitely small ParS

This is the first approach attempted and it highly inspired by [1]. Indeed this way to build the model aims to be a generalization of [1]. ParS is considered to be infinitely small in size but still able to have a finite influence on the system. Therefore the natural modelling choice for such an element is through a Dirac Delta.

$$\delta_{i,0} \to l_{ParB}\delta(x) \tag{4.17}$$

Where the l_{ParS} is the ParB protein footprint. The Master Equation written in this way has been solved (in the next chapters) and also found a limit where the more general equations (4.14) reduce to the equations of [1].

Finally we have:

$$\dot{\vec{\rho}} = D\frac{\partial^2 \vec{\rho}}{\partial x^2} - (W + \delta(x)W_0)\vec{\rho} + \vec{c} \tag{4.18}$$

where:

$$D = \begin{pmatrix} D_o & 0 \\ 0 & D_c \end{pmatrix} \tag{4.19}$$

Up to this day, this model creates some nonphysical behaviours that are not yet understood.

4.4 Global steady state conditions

It is very interesting to study the integral conditions that solution 6.13 has to fulfill in order to be a steady state. In such a state clearly the total number of dimers has to remain constant in time. For this reason the total flux (over all the DNA considered in the domain) of ParB attaching has to equal the one of the dimers detaching:

$$w_{on}\int dx = k_{off}\int dx\,(\rho^o_\infty + \Pi^o(x)) \tag{4.20}$$

Both w_{on} and $k_{off}\rho^o_\infty$ are not space dependent, and they are equal to each other, thus canceling each other out. We are left with:

$$\delta N^o = \int dx \Pi(x) = 0 \tag{4.21}$$

Since this function varies in space we have that this density has to be negative in some parts of the domain. This is acceptable as long as $\rho(x) = \Pi(x) + \rho_\infty \geq 0$. Concerning the opening and closure rates the steady state must fulfill:

$$\int k_{oc}(x) \left(\rho^o_\infty + \Pi^o(x)\right) \, dx = \int k_{co}(x) \left(\rho^c_\infty + \Pi^c(x)\right) \, dx \tag{4.22}$$

the opening and closure rates are vary in space and they have the following form:

$$k_{oc}(x) = k_{oc} + \delta(x)\Delta k_{oc} \tag{4.23}$$

Exploiting the fact that the flat distribution is an equilibrium distribution and also using (**??**):

$$k^S_{oc}\rho^o_\infty + (k^S_{oc} - k_{oc})\Pi^o(x=0) = k_{co}\Delta N^c + k^S_{co}\rho^c_\infty + (k^S_{co} - k_{co})\Pi^c(x=0) \tag{4.24}$$

Chapter 5

Finite size ParS

5.1 The square function ParS

We will consider ParS a small finite region of DNA of size l_s (experiments shows that the sequence is $16bp$ long). This region of DNA has different opening/closure rates.
In particular, for $x \in [-d, +d]$

$$\begin{cases} k^s_{oc} = \gamma_{oc} k_{oc} \\ k^s_{co} = \gamma_{co} k_{co} \end{cases} \tag{5.1}$$

Elsewhere:

$$\begin{cases} k_{oc} \\ k_{co} \end{cases} \tag{5.2}$$

In this model all the other rates (binding, unbinding,..) remain the same regardless if they are in the centromere section or in the rest of the DNA. This choice is justified by the fact that the only difference between the ParS sequence and the rest of DNA that has been observed in experiments is an increase in the closure rate (k_{oc}). Finally we obtain two different equations for the centromere region and for the rest of the DNA, respectively:

$$D\frac{\partial^2 \vec{\rho}}{\partial x^2} - W\vec{\rho} + \phi = 0$$

$$D\frac{\partial^2 \vec{\rho}}{\partial x^2} - W_s\vec{\rho} + \phi = 0$$

For Non Specific DNA we have:

$$\begin{cases} D_o \frac{\partial^2 \rho_o}{\partial x^2} - k_{oc}\rho_o + k_{co}\rho_c - k_{off}\rho_o + \phi = 0 \\ D_c \frac{\partial^2 \rho_c}{\partial x^2} + k_{oc}\rho_o - k_{co}\rho_c = 0 \end{cases}$$

For the ParS interval:

$$\begin{cases} D_o \frac{\partial^2 \rho_o}{\partial x^2} - k^s_{oc}\rho_o + k^s_{co}\rho_c - k_{off}\rho_o + \phi = 0 \\ D_c \frac{\partial^2 \rho_c}{\partial x^2} + k^s_{oc}\rho_o - k^s_{co}\rho_c = 0 \end{cases}$$

The structure of the solutions is clearly the same for both set of equations, indeed the ParS solution is obtained taking the solution of Non Specific DNA and adding the γ multiplying factor. We than perform decompose the solution in a space dependent and in a space constant part:

$$\vec{\rho}(x) = (\vec{\rho}(x) - \vec{\rho}_\infty) + \vec{\rho}_\infty = \vec{\Pi}(x) + \vec{\rho}_\infty \tag{5.3}$$

It's important to notice that the constant part is the solution of the equation deprived of the diffusive term:

$$\vec{\rho}_\infty = W^{-1}\vec{\phi} \tag{5.4}$$

$$\vec{\rho}^s_\infty = W_s^{-1}\vec{\phi} \tag{5.5}$$

is also the equilibrium distribution, which therefore does not dissipate energy. Such decomposition of the solution leads us to an equation for the space dependent part $\Pi(x)$:

$$D\frac{\partial^2 \vec{\Pi}}{\partial x^2} - W\vec{\Pi} = 0 \tag{5.6}$$

This decomposition is performed for both the ParS subset and the Non-Specific DNA. Introducing matrix $B = D^{-1}W$ calling $\vec{v_i}$ its eigenvectors and λ_i its eigenvalues the solution is Imposing the symmetry of the distribution around the origin, the solution can be rewritten as as linear combination of hyperbolic cosines

$$\begin{cases} \vec{\rho} = b_1\vec{v_1}\cosh\sqrt{\lambda_1}x + b_2\vec{v_2}\cosh\sqrt{\lambda_2}x + \vec{\rho}_\infty \\ \vec{\rho^s} = b^s_1\vec{v^s_1}\cosh\sqrt{\lambda^s_1}x + b^s_2\vec{v^s_2}\cosh\sqrt{\lambda^s_2}x + \vec{\rho}^s_\infty \end{cases} \tag{5.7}$$

Where $\vec{c}$ is:

$$\vec{\rho}_\infty = \frac{w_{on}}{k_{off}} \begin{pmatrix} 1 \\ \frac{k_{oc}}{k_{co}} \end{pmatrix} \tag{5.8}$$

And $\vec{c^s}$ is:

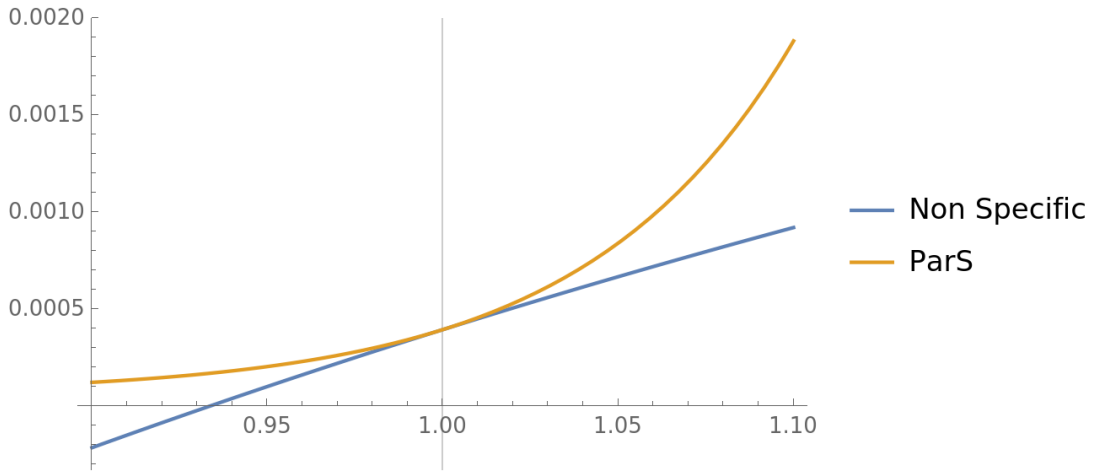
$$\vec{\rho}^s_\infty = \frac{w_{on}}{k_{off}} \begin{pmatrix} 1 \\ \frac{k_{oc}}{k_{co}}\frac{\gamma_{oc}}{\gamma_{co}} \end{pmatrix} \tag{5.9}$$

While the constant parts are the same for the open state, they are not for the closed one. Experiments show a higher closing rate in ParS meaning $\frac{\gamma_{oc}}{\gamma_{co}} > 1$.
Solutions in (5.7) has 4 free parameters. Since the two functions (ρ and ρ_s) describe

one unique density function, we need to impose continuity and differentiability. The four coefficients will be determined "gluing" together the two functions. We have therefore 4 equations for the 4 parameters

$$\begin{cases} \rho_o(x=d) = \rho_o^s(x=d) \\ \rho_c(x=d) = \rho_c^s(x=d) \\ \frac{d}{dx}\rho_o(x=d) = \frac{d}{dx}\rho_o^s(x=d) \\ \frac{d}{dx}\rho_c(x=d) = \frac{d}{dx}\rho_c^s(x=d) \end{cases} \tag{5.10}$$

Figure 5.1: The two solutions are tangent in $x = d/2$, ensuring continuity of both the function and its derivative



These equations have been symbolically solved. These analytical solutions are very long and therefore will not be transcribed in this thesis. One important property of these coefficients is that they are all proportional to $(\gamma_{oc} - \gamma_{co})$ meaning that if the two opposite rates are accelerated by the same factor we have a flat solution.

Figure 5.2: The solution for the open-state dimers, we can obserbe that the open state, close to the ParS region is almost depleted due to the accelerated closure reaction ($\gamma_{oc} > \gamma_{co}$)

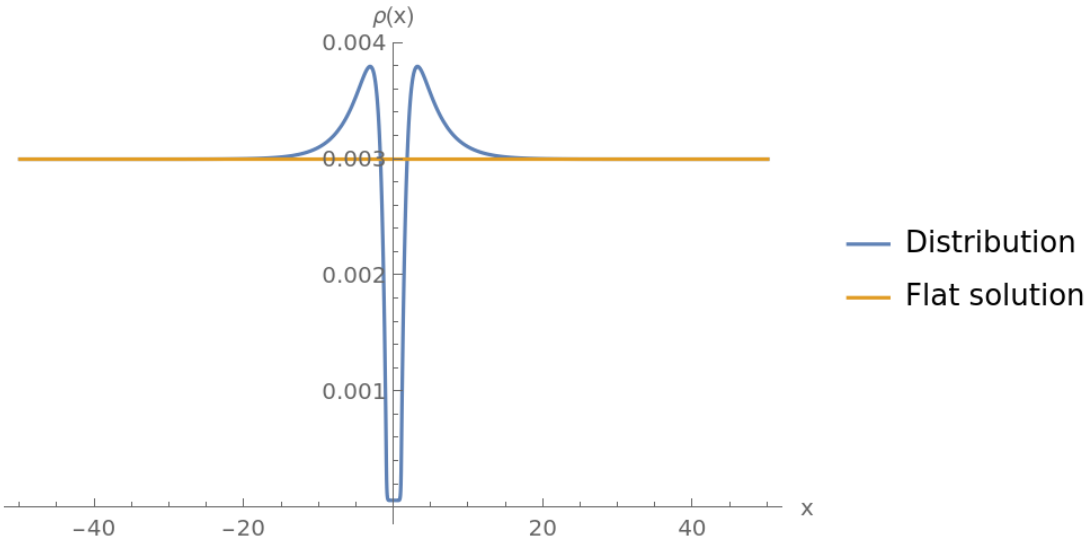


Figure 5.3: The closed-state distribution of dimers is peaked

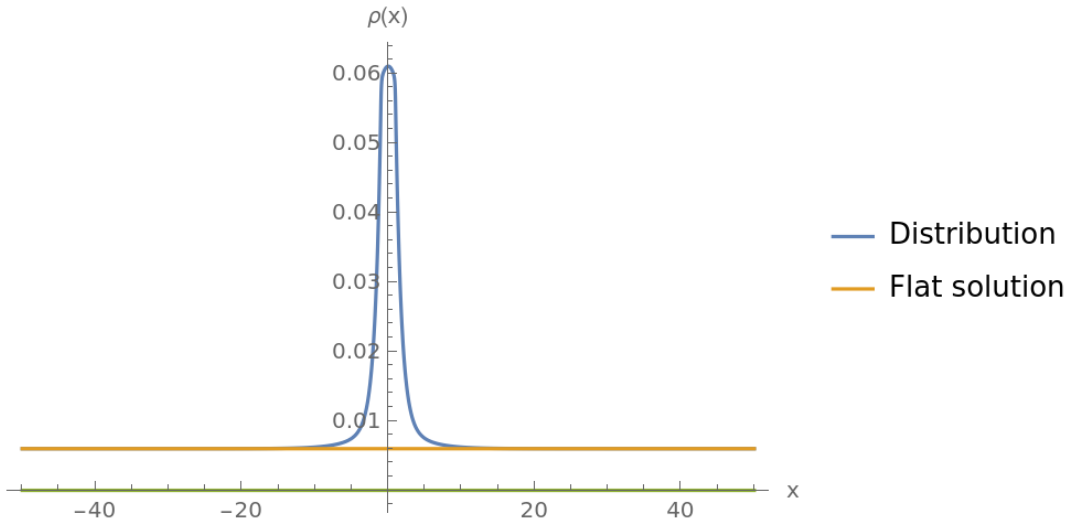
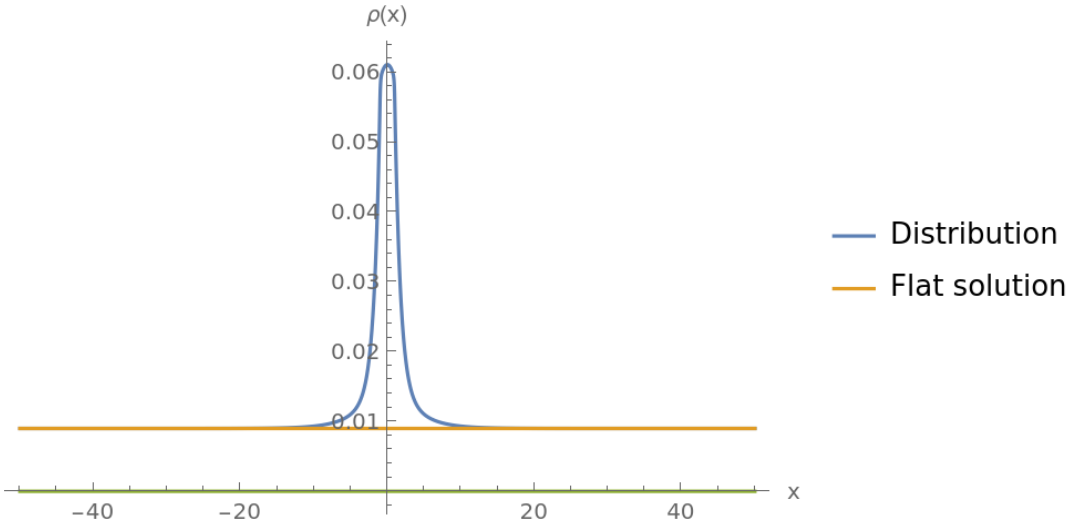


Figure 5.4: This grapsh is the sum of the two previous graphs. This distribution is of great importance because it is the one that can be compared to experiments, since there is no experimental way to distinguish between open and closed configurations when measuring the density. Here the height of the spike is almost 7 times the height of the flat solution baseline.



5.2 Using a finite ParB concentration

So far this model has been worked out in the approximation that the binding of ParB dimers on DNA does not affect the solution concentration. While this could be a reasonable approximation if the concentration is very high, experiments show that this is not the case. In fact it has been observed ([5]) that the number of dimers is around $N = 50/100$. In this chapter is shown that introducing a fixed number of dimers cause the differential equation to become non-local and therefore long-range interaction emerge.

The relatively low total number of dimers per cell can be divided in the ones clamped to DNA and the one that are free in the cell:

$$N = N_{DNA} + N_{sol} \tag{5.11}$$

Once introduced this variables one can see the approximation of the previous section with perspective: there was no fixed N number of dimers, and the number of dimers in solution was considered fixed:

$$\begin{cases} \dot{N}_{DNA} \propto +k_{on} N_{sol} - k_{off} N_{DNA} \\ \dot{N}_{sol} = 0 \end{cases} \tag{5.12}$$

Once we release this assumption the master equation for the number of ParB becomes:

$$\begin{cases} \dot{N}_{DNA} \propto +k_{on}N_{sol} - k_{off}N_{DNA} \\ \dot{N}_{sol} \propto -k_{on}N_{sol} + k_{off}N_{DNA} \end{cases} \tag{5.13}$$

The two differential equation sum to zero, stemming from the conservation (6.21). In particular we have that:

$$N_{DNA} = \int dx \, (\Pi^o(x) + \Pi^c(x) + \rho^o_\infty + \rho^c_\infty) = \delta N + 2L\rho_\infty \tag{5.14}$$

Where $\delta N = \delta N^o + \delta N^c$ is the sum of the integrals over the whole domain respectively of Π^o and Π^c. Another remark has to be done about δN: being the integral of the density over the whole domain, it also includes the ParS region:

$$\delta N = \delta N_{NS} + \delta N_{ParS} \tag{5.15}$$

5.2.1 The w_{on} rate and the constant distribution

In this formulation of the model we need a more accurate definition of the binding rate w_{on}. In fact, in accord with (5.13) we have that the binding rate depends on the dimer-concentration of the solution. Treating this concentration as an homogeneous linear concentration (we consider DNA as mono-dimensional) we have:

$$w_{on} = k_{on}[ParB]_{sol} \tag{5.16}$$

The approximation that makes the concentration of dimers in the solution homogeneous in space is justified by the fact that the diffusion constant D_{sol} is orders of magnitude higher than the diffusion constant D_{DNA} that we considered in the previous section.

$$[ParB]_{sol} = k_{on}\frac{N - N_{DNA}}{2L} = k_{on}\frac{N - \delta N - 2L\rho_\infty}{2L} \tag{5.17}$$

Where 2L is the size of the system. An important characteristic of the equation that is obtained through this definition of $w_{on} = w_{on}(\delta N)$ is that the differential equation for the space dependent function $\Pi(x)$ becomes non local since it depends on its definite integral.

Solving the equation in the ParS-less limit, and for a constant solution, expression (6.14) is found:

$$\vec{\rho}_\infty = \frac{w_{on}}{k_{off}k_{co}} \begin{pmatrix} k_{co} \\ k_{oc} \end{pmatrix} \tag{5.18}$$

ρ_∞ is a function of w_{on} which in turn (6.25) is a function of ρ_∞. Considering the sum of the entries of the vector we have an equation where ρ_∞ appears in both sides:

$$\rho_\infty = \rho^o_\infty + \rho^c_\infty$$

we have a:

$$\rho_\infty = \frac{k_{on}(k_{oc}+k_{co})}{k_{off}k_{co}2L}\left(N - \delta N - \rho_\infty 2L\right) \tag{5.19}$$

defining the constant α,

$$\alpha = \frac{k_{on}}{k_{off}k_{co} + k_{on}(k_{oc}+k_{co})}$$

the solution is:

$$\rho_\infty = \frac{\alpha}{2L}(k_{oc}+k_{co})(N-\delta N) \tag{5.20}$$

the ρ_∞ vector is finally found and it is:

$$\vec{\rho}_\infty = \frac{\alpha}{2L}(N-\delta N)\begin{pmatrix} k_{oc} \\ k_{co} \end{pmatrix} \tag{5.21}$$

at the same way we find $\vec{\rho}^s_\infty$:

$$\vec{\rho}^s_\infty = \frac{\alpha}{2L}(N-\delta N)\begin{pmatrix} k_{oc} \\ k_{co}\frac{\gamma_{oc}}{\gamma_{co}} \end{pmatrix} \tag{5.22}$$

5.2.2 Finding δN

We now proceed to solve the equations and to glue them together (5.10) exactly as in the previous section. The main difference is that, once found the 4 b_i coefficients present in 5.7 they all depend on δN. Recalling its definition:

$$\delta N = \int dx \Pi^o(x) + \int dx \Pi^c(x) \tag{5.23}$$

Also recalling 5.15 we can decompose the integral in the two contribution coming from the two domains:

$$\delta N_{NS} = \int_d^L dx \left(\Pi^o(x) + \Pi^c(x)\right) \tag{5.24}$$

We have to be careful for the integral around the origin. We need to sum the special constant distribution and subtract the NonSpecial one. For the open distribution the integral reads:

$$\delta N^o_{ParS} = \int_0^d dx \left(\Pi^o_s(x) + \rho^s_{\infty,o} - \rho_{\infty,o}\right) \tag{5.25}$$

Recalling that the integral of an hyperbolic cosine is:

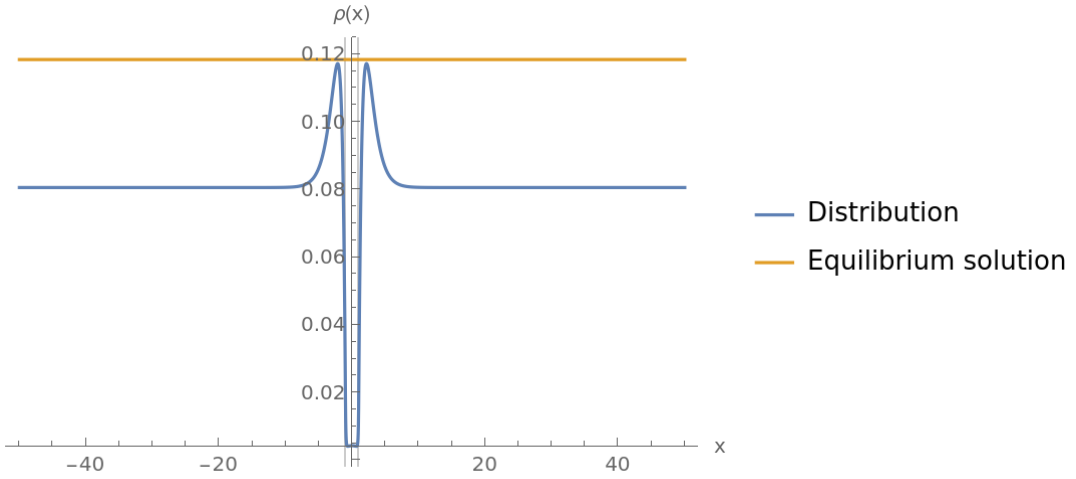
$$\int \cosh\left(\sqrt{\lambda}x\right) dx = \frac{\sinh\left(\sqrt{\lambda}x\right)}{\sqrt{\lambda}} + c \tag{5.26}$$

We can finally sum all these contribution to find the expression of δN which also depends on itself since the expression of the flat distributions ρ_∞ do (5.21). Therefore we get an equation where δN appears on both sides:

$$\delta N = f(\delta N)$$

Solving the equation above for δN we finally find its expression. Substituting this new expression in all the the expression that depend on it, we finally arrive to the solution of this model. The plots of the spatial distributions found are represented below:

Figure 5.5: The solution for the open-state dimers, we can obserbe that the open state, close to the ParS region is almost depleted due to the accelerated closure reaction ($\gamma_{oc} > \gamma_{co}$)



This model succeeds in concentrating the dimers in the non constant solution (lowering the baseline). However experiments [7] show that the peak should orders of magnitude wider than the size of the ParS sequence fig.(5.8).

Figure 5.6: The closed-state distribution of dimers is also peaked here, however the peak, when compared to the baseline of the flat distribution, is considerably higher. This is due to the fact that, having a finite number of dimers, if most of them are clustered around ParS in the space varying part of the solution they can not be in the flat distribution.

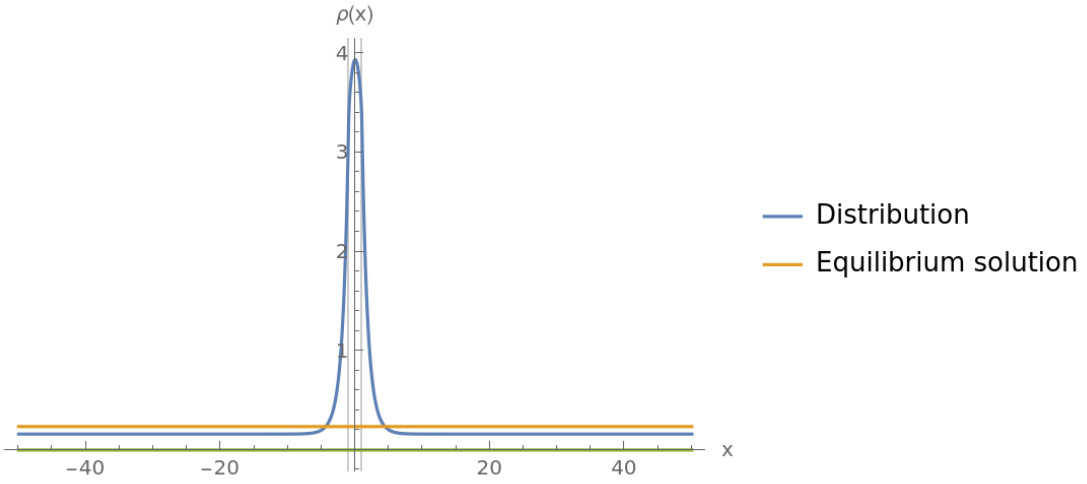


Figure 5.7: This graph is the sum of the two previous graphs. It's very easy to notice that it strongly resembles the closed-state distribution, implying that the letter it is the dominant term among the two. This distribution is of great importance because it is the one that can be compared to experiments, since there is no experimental way to distinguish between open and closed configurations when measuring the density.

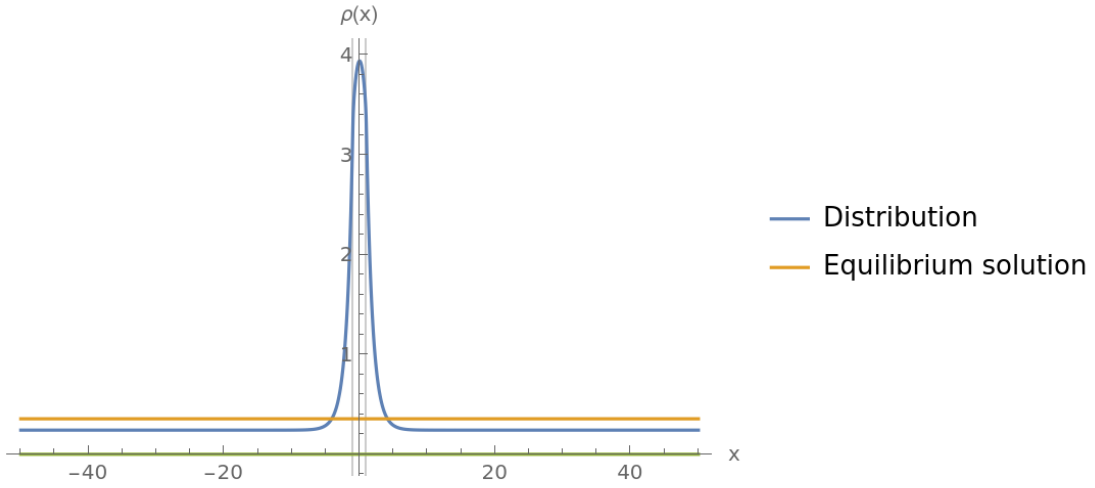
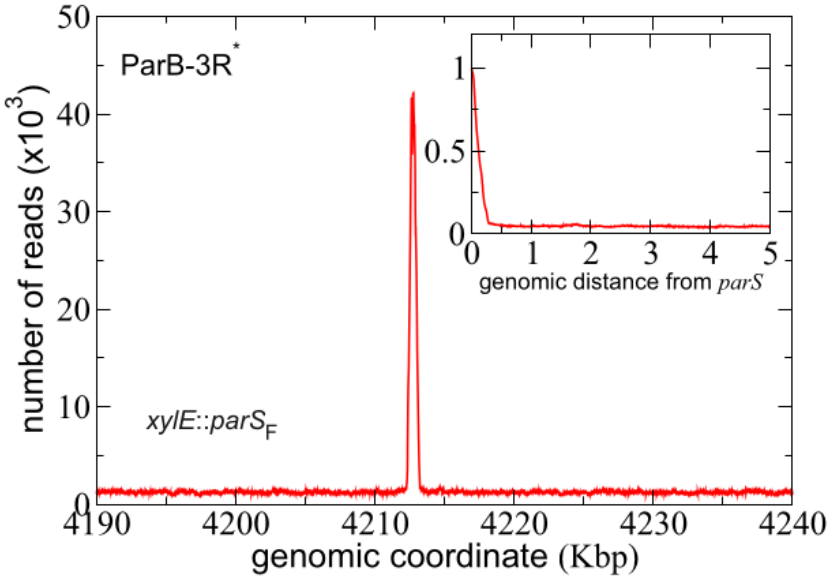


Figure 5.8: [7] This experimental distribution shows that the width of the spike is around 500 bp while the size of ParS is around 20bp



5.3 Maximization of the spike

In the previous section an analytical solution for this system has been found. The great advantage to have an analytical solution (even if its structure is too complicated to be easily studied) is that we can study and plot how some significant quantities vary with the parameters. In particular this section will focus on δN. This quantity is the number of dimers present in the non-flat solution:

$$N_{tot} = N_{DNA} + N_{solution} = \delta N + N_{flat} + N_{solution} \tag{5.27}$$

Experiments show that during chromosome separation almost all dimers cluster around the ParS sequence, meaning that both the solution ($N_{solution}$) and the flat solution ($N_{solution}$) are almost depleted.
Studying how these quantity varies with the parameter is an incredibly challenging: it is a function of 8 variables and navigating such phase space is hard. In order to understand how δN varies along with some of these parameters, we gave values to all of them but on or two, in order to generate meaningful plots.
In particular the focus of this section is on the 2 parameters that have been found influencing the most δN: the two diffusion constants. In particular it has been found that when the closed state constant is much smaller than the open-state high values of δN (meaning that they tend to N_{tot}) are obtained.

Figure 5.9: In this plot all parameters are fixed to the values that generated all the plots above, except the closed-state coefficient of diffusion D_c (re scaled). The =1 vertical line shows the case where $D_o = D_c$. As we can see values of D_o/D_c smaller than 10^{-2} bring almost the entirety of the dimers in the Non-constant solution

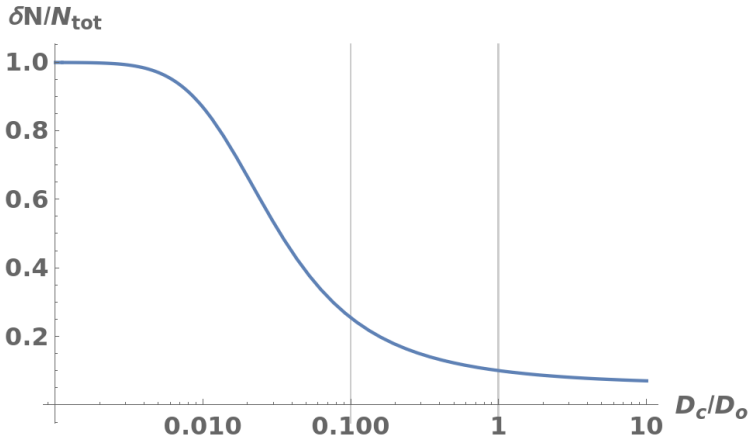


Figure 5.10: Fixing all parameters except both diffusion constants

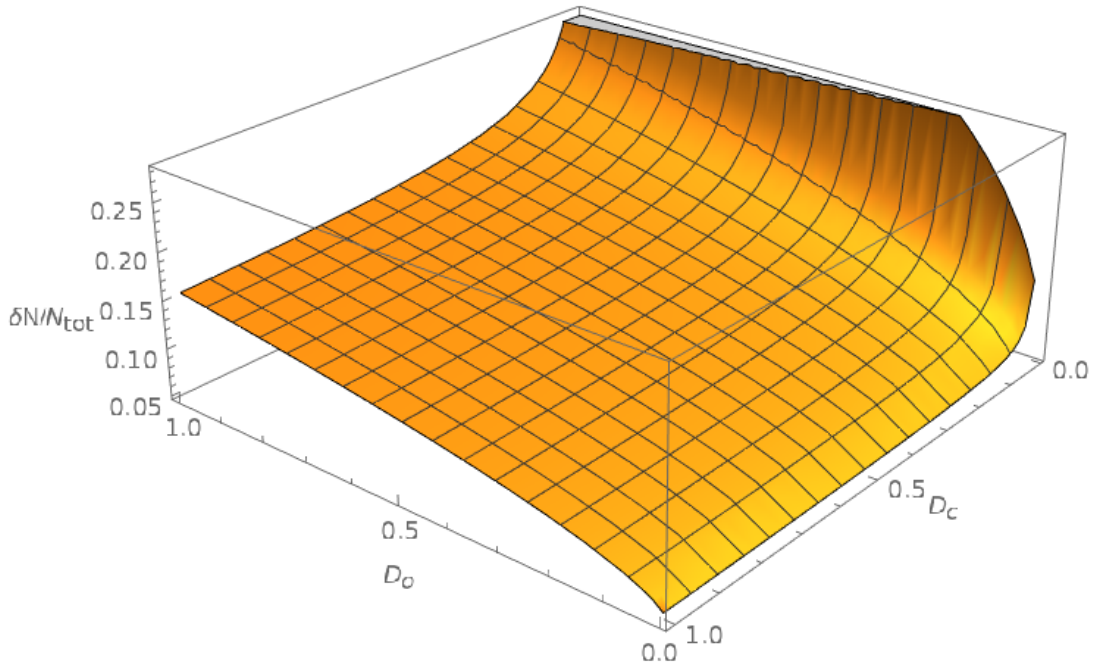
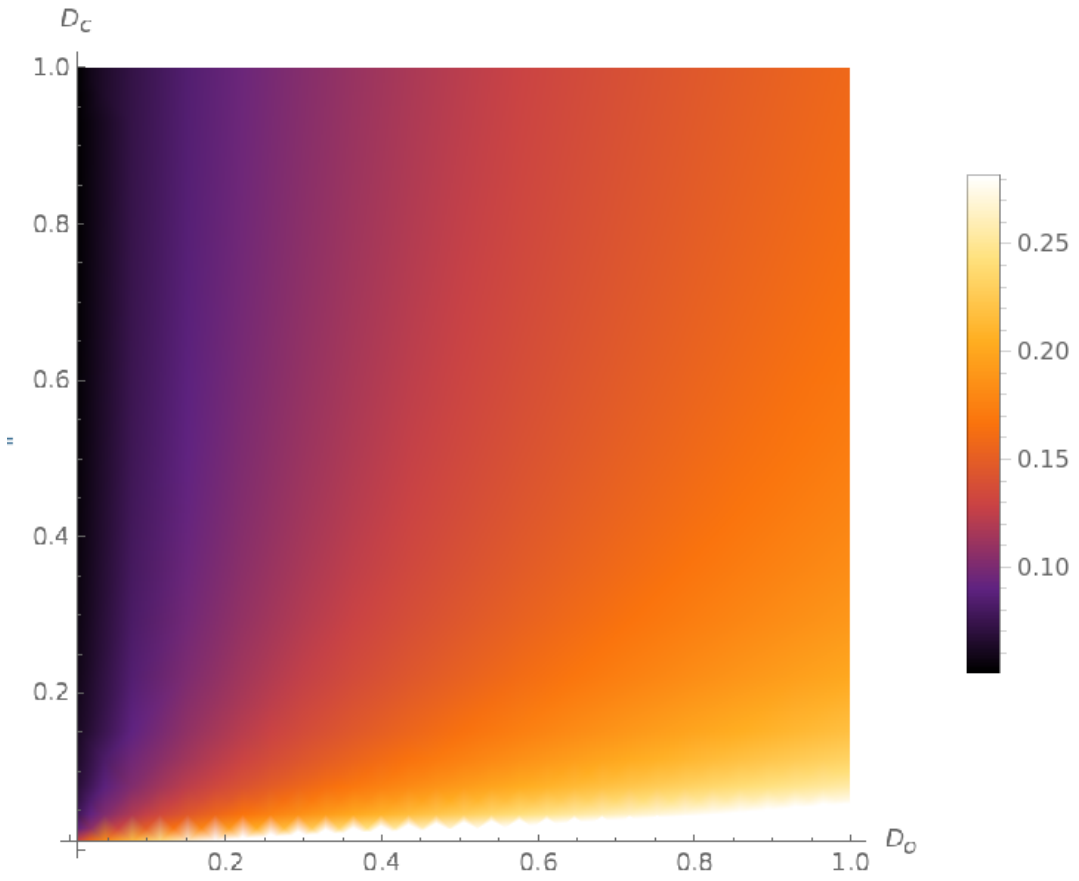


Figure 5.11: Same graph as before but represented through a color map



Chapter 6

Modelling ParS with a Dirac Delta

This first approach has been the first one tried and the one which has been more vastly explored. The next chapter describes an interesting alternative way to consider this problem which finds an interesting connection with the model proposed in [1].

6.1 Solving the equation with diagonalization

Considering the limit for $x \to \infty$ we have that the derivative vanishes, and so does the delta. Setting the time derivative to zero (Steady State) the equation becomes:

$$W\vec{\rho} = \vec{c}$$

$$\rho_\infty = W^{-1}\vec{c} \tag{6.1}$$

which is clearly a constant solution. Decomposing the density in a space dependent term and the former constant one:

$$\rho(x) = (\rho(x) - \rho_\infty) + \rho_\infty = \Pi(x) + \rho_\infty \tag{6.2}$$

The equation simplifies in:

$$D\frac{\partial^2\vec{\Pi}}{\partial x^2} - (W + \delta(x)W_0)\vec{\Pi} - \delta(x)W_0\vec{\rho}_\infty = 0 \tag{6.3}$$

Where the last term could be the analog of the source term in [1].

This system of ODE needs to be eigen-decomposed. Multiplying by D^{-1} and decomposing:

$$D^{-1}W = U\Lambda U^{-1}$$

and defining $\Pi' = U^{-1}\Pi$

$$\frac{\partial^2 \vec{\Pi}'}{\partial x^2} - \Lambda\vec{\Pi}' - \delta(x)U^{-1}D^{-1}W_0(\vec{\Pi} + \vec{\rho_\infty}) = 0$$

Fourier transforming the expression we have:

$$-k^2\vec{\Pi}'(k) - \Lambda\vec{\Pi}'(k) - U^{-1}D^{-1}W_0(\vec{\Pi}(x=0) + \vec{\rho}_\infty) = 0$$

We finally get to the Fourier transform of the solution:

$$\vec{\Pi}'(k) = -(k^2 I + \Lambda)^{-1}U^{-1}D^{-1}W_0(\vec{\Pi}(x=0) + \vec{\rho}_\infty) \tag{6.4}$$

Focusing on the k-dependent diagonal matrix:

$$(k^2 I + \Lambda)^{-1} = \begin{pmatrix} k^2+\lambda_1 & 0 \\ 0 & k^2+\lambda_2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{k^2+\lambda_1} & 0 \\ 0 & \frac{1}{k^2+\lambda_2} \end{pmatrix} \tag{6.5}$$

In order to get the desired expression in the position space:

$$\int e^{ikx}(k^2 I + \Lambda)^{-1}\,dk = \begin{pmatrix} \frac{1}{2\sqrt{\lambda_1}}e^{-\sqrt{\lambda_1}|x|} & 0 \\ 0 & \frac{1}{2\sqrt{\lambda_2}}e^{-\sqrt{\lambda_2}|x|} \end{pmatrix} = \frac{1}{2}Q(x) \tag{6.6}$$

Remembering that $\Pi' = U^{-1}\Pi$ and multiplying by U we finally get:

$$\vec{\Pi}(x) = -\frac{1}{2}UQ(x)U^{-1}D^{-1}W_0(\vec{\Pi}(x=0) + \vec{\rho}_\infty) \tag{6.7}$$

The structure of this solution is the linear combination of the two exponentials. Defining

$$A(x) = \frac{1}{2}UQ(x)U^{-1}D^{-1} \tag{6.8}$$

$$\vec{\Pi}(x) = -A(x)W_0(\vec{\Pi}(0) + \vec{\rho}_\infty) \tag{6.9}$$

and setting x to zero and solving for $\Pi(x=0)$ we find:

$$\Pi(0) = -(I + A(0)W_0)^{-1}A(0)W_0\rho_\infty \tag{6.10}$$

Substituting 6.10 in 6.9 and collecting terms, we get:

$$\vec{\Pi}(x) = A(x)[W_0(I + A(0)W_0)^{-1}A(0) - I]W_0\rho_\infty \tag{6.11}$$

Condensing all matrices together we have:

$$\vec{\Pi}(x) = T(x)\vec{\rho}_\infty \tag{6.12}$$

6.1.1 The solution

The solution is a vector, where both components are linear combination of the same exponentials:

$$\vec{\rho} = \begin{pmatrix} c_1^o e^{-\sqrt{\lambda_1}|x|} + c_2^o e^{-\sqrt{\lambda_2}|x|} + \rho_{\infty o} \\ c_1^c e^{-\sqrt{\lambda_1}|x|} + c_2^c e^{-\sqrt{\lambda_2}|x|} + \rho_{\infty c} \end{pmatrix} \tag{6.13}$$

The model developed so far is very general, and as a price for its generality it has many parameters. The higher number of parameters make the analytical solution hard to interpret.
For this reason, a simpler model is sought. Many simplifications inspired by experiments can be made.
First of all we will consider that only dimer in the open state can attach and detach from DNA. This corresponds to setting $k_{off}^c = k_{on}^c = 0$. Under this assumption the expression of ρ_∞ is:

$$\rho_\infty = \frac{w_{on}}{k_{off} k_{co}^{NS}} \begin{pmatrix} k_{co}^{NS} \\ k_{oc}^{NS} \end{pmatrix} \tag{6.14}$$

6.1.2 The role of the ParS locus and the dissipation of energy

Probably the most crucial element of this model is the ParS site. Various experiments show large clusters of dimers around it. It is evident that this locus (positioned in the origin of the model) interacts with the dimers. Experiments show ([5]) that this DNA segment is able to act as a catalyst for the opening and closure reactions.

In order to understand the role of this gene in the model we will focus on the matrix vector product:

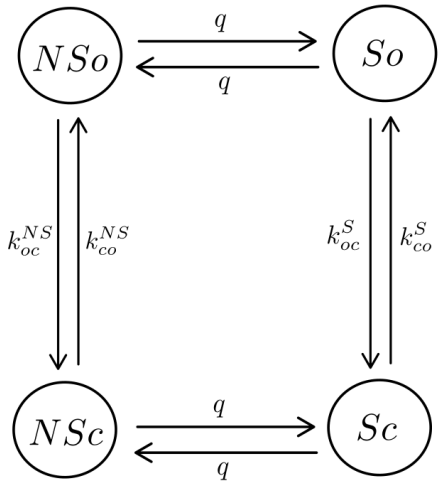
$$W_0 \rho_\infty = \begin{pmatrix} -(k_{oc}^S - k_{oc}^{NS}) & +(k_{co}^S - k_{co}^{NS}) \\ +(k_{oc}^S - k_{oc}^{NS}) & -(k_{co}^S - k_{co}^{NS}) \end{pmatrix} \frac{w_{on}}{k_{off} k_{co}^{NS}} \begin{pmatrix} k_{co}^{NS} \\ k_{oc}^{NS} \end{pmatrix} \tag{6.15}$$

If we decompose the matrix: $W_0 = M^S - M^{NS}$ we get that $M^{NS}\rho_\infty = 0$ which implies:

$$W_0 \rho_\infty = W^S \rho_\infty = \frac{w_{on}}{k_{off} k_{co}^{NS}} \left(+k_{oc}^S k_{co}^{NS} - k_{co}^S k_{oc}^{NS} \right) \begin{pmatrix} -1 \\ +1 \end{pmatrix} \tag{6.16}$$

The expression that appears in the parenthesis has an important physical interpretation: it is the product of the rates (of the reaction graph below) in one direction minus the ones in the opposite (the hopping q rates cancel out):

Figure 6.1: Reaction graph for ParS locus and one of the two neighbor sites



Using relation (1.32) we rewrite:

$$\left(+k_{oc}^{S}k_{co}^{NS} - k_{co}^{S}k_{oc}^{NS}\right) = \left(k_{co}^{S}k_{oc}^{NS}\right)\left(e^{-\frac{F}{k_bT}} - 1\right) \tag{6.17}$$

Finally we can write:

$$\Pi(x) \propto \left(e^{-\frac{\Delta G}{k_bT}} - 1\right) \tag{6.18}$$

This last proportionality has a profound physical meaning: the non constant solution can exist only if the system is out of equilibrium and thus dissipates free energy. The cycle that is linked to the energy consumption of this model is the one of fig.6.1.

6.1.3 The 2 branches

Recalling chapter 3, the opening and closure rates are just the sum of the two branches D and T.

$$\begin{cases} k_{oc} = k_{oc,D} + k_{oc,T} \\ k_{oc}^{S} = \gamma_D k_{oc,D} + \gamma_T k_{oc,T} \\ k_{co} = k_{co,D} + k_{co,T} \\ k_{co}^{S} = \gamma_D k_{co,D} + \gamma_T k_{co,T} \end{cases} \tag{6.19}$$

Substituting this relations in the model we obtain that

$$\Pi(x) \propto \frac{w_{\text{on}}\left(\gamma_D - \gamma_T\right)\left(k_{\text{co},T}k_{\text{oc},D} - k_{\text{co},D}k_{\text{oc},T}\right)}{k_{\text{off}}\left(k_{\text{co},D} + k_{\text{co},T}\right)} \tag{6.20}$$

The former relation gives a more detailed description of the out of equilibrium properties of this system. The first parenthesis at the numerator tells us that in order to have a non-flat solution two different scaling constants are needed for the D and T branch. The second parenthesis it has the same structure of (6.17) but there is a crucial difference: all the rates that appear are Non Specific (before it was half specific, and half non specific). This implies that there will be a thermodynamical force dissipating energy all over DNA. It's really important to notice that the coarse-graining technique developed in chapter 4 does conserve this particular thermodyanamical force (3.25).

6.2 Finite ParB concentration

So far this model has been worked out in the approximation that the binding of ParB dimers on DNA can not affect the the solution concentration. While this could be a reasonable approximation if the concentration is very high, this is not the case. In fact it has been observed [5] that the number of dimers is around $N = 60$.

$$N = N_{DNA} + N_{sol} \tag{6.21}$$

Once introduced this variables one can define the previous approximation in the following way:

$$\begin{cases} \dot{N}_{DNA} \propto +k_{on}N_{sol} - k_{off}N_{DNA} \\ \dot{N}_{sol} = 0 \end{cases} \tag{6.22}$$

where L is the length of the system. This approximation clearly does not obey relation (6.21) if N is fixed.
Once we release this assumption the master equation for the number of ParB becomes:

$$\begin{cases} \dot{N}_{DNA} \propto +k_{on}N_{sol} - k_{off}N_{DNA} \\ \dot{N}_{sol} \propto -k_{on}N_{sol} + k_{off}N_{DNA} \end{cases} \tag{6.23}$$

In particular we have that:

$$N_{DNA} = \int dx \, (\Pi^o(x) + \Pi^c(x) + \rho^o_\infty + \rho^c_\infty) = \delta N + L\rho_\infty \tag{6.24}$$

Where $\delta N = \delta N^o + \delta N^c$ is the sum of the integrals over the whole domain respectively of Π^o and Π^c. In the differential equation for Π the term $w_{on} = k_{on}[ParB]_{sol}$ (we consider the concentration as linear concentration).

$$w_{on} = k_{on}[ParB]_{sol} = k_{on}\frac{N - N_{DNA}}{L} = k_{on}\frac{N - \delta N - L\rho_\infty}{L} \tag{6.25}$$

The equation for the space dependent part $\Pi(x)$ depends on its definite integral δN over the whole domain.

6.2.1 The equation for ρ_∞

Solving the equation in the ParS-less limit, and for a constant solution, expression (6.14) is found:

$$\vec{\rho}_\infty = \frac{w_{on}}{k_{off}k^{NS}_{co}} \begin{pmatrix} k^{NS}_{co} \\ k^{NS}_{oc} \end{pmatrix} \tag{6.26}$$

ρ_∞ is a function of w_{on} which in turn (6.25) is a function of ρ_∞. Considering the sum of the entries of the vector we have an equation where ρ_∞ appears in both sides:

$$\rho_\infty = \rho_\infty^o + \rho_\infty^c$$

we have a:

$$\rho_\infty = \frac{k_{on}(k_{oc} + k_{co})}{k_{off}k_{co}L}\left(N - \delta N - \rho_\infty L\right) \tag{6.27}$$

defining the constant α,

$$\alpha = \frac{k_{on}}{k_{off}k_{co} + k_{on}(k_{oc} + k_{co})}$$

the solution is:

$$\rho_\infty = \frac{\alpha}{L}(k_{oc} + k_{co})(N - \delta N) \tag{6.28}$$

the ρ_∞ vector is finally found and it is a function of :

$$\vec{\rho_\infty} = \frac{\alpha}{L}(N - \delta N)\begin{pmatrix} k_{oc} \\ k_{co} \end{pmatrix} \tag{6.29}$$

6.2.2 Solving the equation

Recalling structure of the solution (6.9):

$$\vec{\Pi(x)} = -A(x)W_0\left(\vec{\Pi}(0) + \vec{\rho}_\infty(\delta N)\right) \tag{6.30}$$

This solution depends both on it's value at the origin and its definite integral. Integrating both sides we have:

$$\begin{pmatrix} \delta N^o \\ \delta N^c \end{pmatrix} = -BW_0\left(\vec{\Pi}(0) + \vec{\rho}_\infty(\delta N)\right) \tag{6.31}$$

where B is the integral of $A(x)$. The integral of the matrix is trivial since it's the integral of exponentials. L is considered big enough to justify the approximation the integral on domain $[-\frac{L}{2}, \frac{L}{2}]$ with $[-\infty, +\infty]$. Summing the two equation an equation in δN is obtained and it will be solved in such variable:

$$\delta N = -\sum_{i=o,c}\left(BW_0\vec{\Pi}(0)\right)_i - \sum_{i=o,c}\left(BW_0\vec{\rho_\infty}(\delta N)\right)_i \tag{6.32}$$

Solving for δN we find:

$$\delta N = \frac{\sum_{i=o,c}\left(BW_0\vec{\Pi}(0)\right)_i}{1 - \frac{\alpha}{L}\sum_{i=o,c}\left(BW_0\vec{k}\right)_i} \tag{6.33}$$

We finally found the expression of δN as a function of $\Pi(0)$. Substituting in ρ_∞ in turn it becomes dependent solely on $\Pi(0)$. Setting x=0 we we find a generalization of equation (6.10) where now the only unknown is $\Pi(0)$. Solving the equation for $\Pi(0)$ finally solves the system.

6.3 Sum and current equation

The aim of this thesis is to develop a model whose solution is as close as possible to the density function observed in experiments. For this reason the 5-state model was developed. Seeking for a model which can be more easily treated analytically we wrote down the equations of the 3-state systems, and derived the rates with the Coarse-graining procedure described in chapter 4. In this chapter the same model is solved in a slightly different way: rather than writing the equations for the open and closed state density distributions, we will combine them to find the equation for the sum of the two distributions and the current between the them (which is a weighted difference). These two quantities have both a physical interpretation. The goal of this chapter is to find an equation for the sum-distribution that depends only on itself and not on both itself and the current-distribution. The equation found in this way is a generalization of the phenomenological one present in [1].

In order to highlight some important properties of the 3-state system we will perform a change of variables. We will pass from the open and closed state densities to:

$$\begin{cases} \rho = \rho^o + \rho^c \\ \phi = \frac{k_{oc}\rho^o - k_{co}\rho^c}{k_{oc}+k_{co}} = \frac{k_{oc}\rho^o - k_{co}\rho^c}{K} \end{cases} \tag{6.34}$$

The first density is the sum of the two, the second one is the current between open and closed states divided by the sum of the rates. For sake of simplicity we will define $K = k_{oc} + k_{co}$.
The opposite change of variables is given by:

$$\begin{cases} \rho^o = \frac{k_{co}}{K}\rho + \phi \\ \rho^c = \frac{k_{oc}}{K}\rho - \phi \end{cases} \tag{6.35}$$

Starting from:

$$\begin{cases} \dot{\rho^o} = D_o\partial_x^2\rho^o + k_{co}\rho^c - k_{oc}\rho^o - k_{off}\rho^o + w_{on} + \delta(x)\left(+\Delta k_{co}\rho^c - \Delta k_{oc}\rho^o\right) \\ \dot{\rho^c} = D_c\partial_x^2\rho^c - k_{co}\rho^c + k_{oc}\rho^o + \delta(x)\left(-\Delta k_{co}\rho^c + \Delta k_{oc}\rho^o\right) \end{cases} \tag{6.36}$$

At first we will consider the sum equation ($\dot{\rho^o} + \dot{\rho^c}$). Clearly this equation does not depend on the opening-closure rates because the cancel out. Performing the

change of variables relations (6.35) the and defining some constants we get:

$$D\partial_x^2\rho + \Delta D\partial_x^2\phi + w_{on} - k_{off}(\frac{k_{co}}{K}\rho + \phi) = 0 \tag{6.37}$$

Where:

$$\begin{cases} D = \frac{k_{co}D_o + k_{oc}D_c}{K} \\ \Delta D = D_o - D_c \\ D^* = \frac{k_{co}D_c + k_{oc}D_o}{K} \end{cases}$$

Considering the current equation: $\frac{k_{oc}\dot{\rho^o} - k_{co}\dot{\rho^c}}{K}$ and performing the same change of variables we get:

$$\frac{k_{oc}k_{co}}{K^2}\Delta D\partial_x^2\rho + D^*\partial_x^2\phi - \frac{k_{oc}k_{co}}{K^2}\rho - \left(K + \frac{k_{oc}k_{off}}{K}\right)\phi +$$

$$+\delta(x)\left(\frac{-\Delta k_{oc}k_{co} + \Delta k_{co}k_{oc}}{K}\right)\rho + -\delta(x)(\Delta k_{oc} + \Delta k_{co})\phi = 0 \tag{6.38}$$

6.3.1 Solving the sum equation

In order to solve this system of equations we will proceed again decomposing the space-dependent part from the constant part of the solution.

$$\begin{cases} \rho = \delta\rho(x) + \rho_\infty \\ \phi = \delta\phi(x) + \phi_\infty \end{cases} \tag{6.39}$$

In the the previous chapter (in the open-close basis) it was found that the constant solution was:

$$\rho_\infty^{old\_coordinates} = \frac{w_{on}}{k_{off}k_{co}}\begin{pmatrix} k_{co} \\ k_{oc} \end{pmatrix}$$

Finding the sum and current associated to this distribution we find:

$$\rho_\infty = \begin{pmatrix} \frac{w_{on}K}{k_{off}k_{co}} \\ 0 \end{pmatrix} \tag{6.40}$$

Where it is found that $\phi_\infty = 0$ thus implying that $\delta\phi = \phi$. This vector solves the system when both diffusive terms and Dirac delta terms are set to zero. The sum equation becomes (from now on ρ and R will be the sum and current of the):

$$D\partial_x^2\delta\rho + \Delta D\partial_x^2\phi - k_{off}(\frac{k_{co}}{K}\delta\rho + \phi) = 0 \tag{6.41}$$

While the current one:

$$\frac{k_{oc}k_{co}}{K^2}\Delta D\partial_x^2\delta\rho + D^*\partial_x^2\phi - \left(K + \frac{k_{oc}k_{off}}{K}\right)\phi + \delta(x)\left(\frac{-\Delta k_{oc}k_{co} + \Delta k_{co}k_{oc}}{K}\right)\delta\rho +$$

$$-\frac{k_{oc}k_{co}}{K^2}\delta\rho + \delta(x)(\rho_\infty^o + \rho_\infty^c)\left(\frac{-\Delta k_{oc}k_{co} + \Delta k_{co}k_{oc}}{K}\right) - \delta(x)(\Delta k_{oc} + \Delta k_{co})\phi = 0 \quad (6.42)$$

The w_{on} term disappeared from the sum equation, and an additional term appeared in the current equation. Fourier transforming both equations:

$$-q^2 D\delta\rho(q) - q^2\Delta D\phi(q) - k_{off}\left(\frac{k_{co}}{K}\delta\rho(q) + \phi(q)\right) = 0$$

and

$$-q^2\frac{k_{oc}k_{co}}{K^2}\Delta D\delta\rho(q) - q^2 D^*\phi - \left(K + \frac{k_{oc}k_{off}}{K}\right)\phi(q) + \left(\frac{-\Delta k_{oc}k_{co} + \Delta k_{co}k_{oc}}{K}\right)\delta\rho(x=0) +$$

$$-\frac{k_{oc}k_{co}}{K^2}\delta\rho(q) + (\rho_\infty^o + \rho_\infty^c)\left(\frac{-\Delta k_{oc}k_{co} + \Delta k_{co}k_{oc}}{K}\right) - (\Delta k_{oc} + \Delta k_{co})\phi(x=0) = 0 \quad (6.43)$$

Finding the expression of $\phi(q)$ from (6.3.1):

$$\phi(q) = \frac{-k_{\mathrm{co}}k_{\mathrm{off}}\delta\rho(q) - \mathrm{D}Kq^2\delta\rho(q)}{K\left(k_{\mathrm{off}} + \Delta\mathrm{D}q^2\right)} \quad (6.44)$$

Substituting this expression in the current equation we find an equation that now only depends on $\rho(q)$. Solving for $\rho(q)$ and arranging terms we get:

$$\frac{K^2\left(k_{\mathrm{off}} + \Delta\mathrm{D}q^2\right)\left(-\frac{\rho_\infty(\Delta k_{\mathrm{co}}k_{\mathrm{oc}} - k_{\mathrm{co}}\Delta k_{\mathrm{oc}})}{K} - \frac{\rho_0(\Delta k_{\mathrm{co}}k_{\mathrm{oc}} - k_{\mathrm{co}}\Delta k_{\mathrm{oc}})}{K} + \phi_0\left(\Delta k_{\mathrm{co}} + \Delta k_{\mathrm{oc}}\right)\right)}{q^2\left(\mathrm{D}^*Kk_{\mathrm{co}}k_{\mathrm{off}} - 2\Delta\mathrm{D}k_{\mathrm{co}}k_{\mathrm{oc}}k_{\mathrm{off}} + \mathrm{D}Kk_{\mathrm{oc}}k_{\mathrm{off}} + \mathrm{D}K^3\right) + q^4\left(\mathrm{D}\mathrm{D}^*K^2 - \Delta\mathrm{D}^2k_{\mathrm{co}}k_{\mathrm{oc}}\right) + K^2k_{\mathrm{co}}k_{\mathrm{off}}} \quad (6.45)$$

This long expression has the same functional form of:

$$\delta\rho(q) = \frac{dq^2 + g}{aq^4 + bq^2 + c} \quad (6.46)$$

The roots of the second order equation associated to the denominator (after performing a change of variables $y = q^2$) are the inverse of the length scales found in chapter 5 ($\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$).

6.3.2 The sum equation in Real Space (structure)

In order to find the sum equation in real space we will start from the expression of $\phi(q)$ found by solving the current equation (6.3.1):

$$-\frac{\Delta \mathrm{D} q^2 k_{\mathrm{co}} k_{\mathrm{oc}} \delta\rho(q)}{K\left(\mathrm{D}^* K q^2 + k_{\mathrm{oc}} k_{\mathrm{off}} + K^2\right)} + \frac{\rho_\infty \left(K \Delta \mathrm{k}_{\mathrm{co}} k_{\mathrm{oc}} - K k_{\mathrm{co}} \Delta \mathrm{k}_{\mathrm{oc}}\right)}{K\left(\mathrm{D}^* K q^2 + k_{\mathrm{oc}} k_{\mathrm{off}} + K^2\right)} + \frac{\delta\rho_0 \left(K \Delta \mathrm{k}_{\mathrm{co}} k_{\mathrm{oc}} - K k_{\mathrm{co}} \Delta \mathrm{k}_{\mathrm{oc}}\right)}{K\left(\mathrm{D}^* K q^2 + k_{\mathrm{oc}} k_{\mathrm{off}} + K^2\right)} +$$

$$-\frac{k_{\mathrm{co}} k_{\mathrm{oc}} k_{\mathrm{off}} \delta\rho(q)}{K\left(\mathrm{D}^* K q^2 + k_{\mathrm{oc}} k_{\mathrm{off}} + K^2\right)} + \frac{\phi_0 \left(-K^2 \Delta \mathrm{k}_{\mathrm{oc}} - K^2 \Delta \mathrm{k}_{\mathrm{co}}\right)}{K\left(\mathrm{D}^* K q^2 + k_{\mathrm{oc}} k_{\mathrm{off}} + K^2\right)}$$

The structure of $\phi(q)$ is:

$$\phi(q) = \frac{1}{a + q^2}\left(c_{R_0} + c_{\rho_0} + c_{\rho_\infty} + c_0 \delta\rho(q) + c_2 q^2 \delta\rho(q)\right) \tag{6.47}$$

With:

$$a = \frac{k_{oc} k_{off} + K^2}{D^* K} = \frac{k_{oc} k_{off} + K^2}{k_{co} D_c + k_{oc} D_o} \tag{6.48}$$

The function that multiplies the parenthesis is a Lorentzian and its Inverse Transform is:

$$\mathcal{F}^{-1}\left(\frac{1}{a + q^2}\right) = \frac{1}{2\sqrt{a}} e^{-\sqrt{a}|x|} = \Lambda(x)$$

The last term of (6.47) is

$$\mathcal{F}^{-1}\left(\frac{q^2}{a + q^2}\right) = \delta(x) - \frac{1}{2}\sqrt{a} e^{-\sqrt{a}|x|} = \delta(x) - a\Lambda(x) \tag{6.49}$$

Taking the inverse Fourier transform of (6.47) we get:

$$\phi(x) = (c_{R_0} + c_{\rho_0} + c_{\rho_\infty})\Lambda(x) + c_0 \mathcal{F}^{-1}\left(\Lambda(q)\rho(q)\right) + c_2 \mathcal{F}^{-1}\left(\frac{q^2}{a + q^2}\rho(q)\right)$$

Using the Convolution theorem for inverse Fourier transform, and computing the convolution with the Dirac's delta, and rearranging terms:

$$\phi(x) = (c_{R_0} + c_{\rho_0} + c_{\rho_\infty})\Lambda(x) + (c_0 - a c_2)\left(\Lambda * \delta\rho\right)(x) + c_2 \delta\rho(x) \tag{6.50}$$

When we substitute the space dependent expression of (6.50) in (6.41) we finally obtain a non-local differential equation since the convolution operator is non-local.

$$D\partial_x^2\delta\rho + b_1\partial_x^2\left(\Lambda * \delta\rho\right)(x) + b_2\delta\rho + b_3\left(\Lambda * \delta\rho\right)(x) + c(\phi_0, \delta\rho_0, \rho_\infty)\left(a_1\Lambda + a_2\delta(x)\right) = 0 \tag{6.51}$$

Notice that the second derivative of Λ generated a Dirac delta term:

$$\partial_x^2\Lambda(x) = a\Lambda(x) - 2\sqrt{a}\delta(x)\Lambda(x)$$

This is a remarkable property. Summarizing what happened, we formulated a model where there was not an explicit source of dimers (only a region of DNA where some reactions where catalysed). The manipulation of the equations that leads to find an equation that depends only on the sum of the distributions generated a pure source term that was not present before. Equation (6.51) is a generalization of the diffusion equation (4.11) proposed in [1].

Chapter 7

Conclusions and the future the project

ParABS systems drive chromosome separation in most species of bacteria. Understanding the interaction between two of these components and in particular having a physical model for the spatial distribution of ParB dimers on DNA is therefore of great importance. The continuous model proposed in this thesis aims to reproduce the experimental ParB distribution found in [7]. The solution of the Master Equation found in chapter 5 recreates the correct shape but it is too peaked around the ParS gene. The model proposed in chapter 6 has also the correct spiked shape, however it creates some nonphysical behaviours that are still not understood. Fixing both models' flaws will be the first next step of this project. The probability distributions found will than be compared with each other through a Kullback-Leibler divergence [8] [9] in order to have an estimate about the energy dissipation of the Non-equilibrium steady state.

In section 6.2 the master equation has been manipulated in a way that connects this model with the one presented in [1] which partially succeeds in recreating the correct shape of the distribution however ignoring many physical features of the system. Many aspects are still to be studied. One of these is the suspected attractive interaction between open dimers [10] which permits the diffusion of the dimers even in the presence of DNA roadblocks. Another interesting concept that could be investigated and applied to this system is the one of ultra-affinity [11]: the phenomenon that enhances binding through energy consumption.
Much work has been done on this project and much more is needed. However the prospect of publishing (with or without the addition of in vitro experiments) is concrete.

Bibliography

[1] Jean-Charles Walter, Jérôme Rech, Nils-Ole Walliser, Jérôme Dorignac, Fred Geniet, John Palmeri, Andrea Parmeggiani, and Jean-Yves Bouet. «Physical Modeling of a Sliding Clamp Mechanism for the Spreading of ParB at Short Genomic Distance from Bacterial Centromere Sites». In: *SSRN Electronic Journal* 23 (Dec. 2020). DOI: 10.2139/ssrn.3673603 (cit. on pp. i, 15, 24, 26, 41, 48, 52, 53).

[2] Haye Hinrichsen, Christian Gogolin, and Peter Janotta. «Non-equilibrium Dynamics, Thermalization and Entropy Production». In: *Journal of Physics Conference Series* 297 (Feb. 2011). DOI: 10.1088/1742-6596/297/1/012011 (cit. on p. 1).

[3] J. Schnakenberg. «Network theory of microscopic and macroscopic behavior of master equation systems». In: *Rev. Mod. Phys.* 48 (4 Oct. 1976), pp. 571–585. DOI: 10.1103/RevModPhys.48.571. URL: https://link.aps.org/doi/10.1103/RevModPhys.48.571 (cit. on p. 5).

[4] Herbert B Callen. *Thermodynamics and an introduction to thermostatistics; 2nd ed.* New York, NY: Wiley, 1985. URL: https://cds.cern.ch/record/450289 (cit. on p. 6).

[5] Hammam Antar, Young-Min Soh, Stefano Zamuner, Florian Bock, Anna Anchimiuk, Paolo De Los Rios, and Stephan Gruber. «Relief of ParB autoinhibition by parS DNA catalysis and recycling of ParB by CTP hydrolysis promote bacterial centromere assembly». In: *Science advances* 7 (Oct. 2021), eabj2854. DOI: 10.1126/sciadv.abj2854 (cit. on pp. 9, 10, 12, 15, 16, 20, 32, 43, 46).

[6] Adam Jalal and Tung Le. «Bacterial chromosome segregation by the ParABS system». In: *Open Biology* 10 (June 2020), p. 200097. DOI: 10.1098/rsob.200097 (cit. on p. 21).

[7] Aurore Sanchez, Diego I. Cattoni, Jean-Charles Walter, Jérôme Rech, Andrea Parmeggiani, Marcelo Nollmann, and Jean-Yves Bouet. «Stochastic Self-Assembly of ParB Proteins Builds the Bacterial DNA Segregation Apparatus».

In: *Cell Systems* 1.2 (2015), pp. 163–173. ISSN: 2405-4712. DOI: https://doi.org/10.1016/j.cels.2015.07.013. URL: https://www.sciencedirect.com/science/article/pii/S2405471215000575 (cit. on pp. 24, 35, 37, 53).

[8] S. Kullback and R. A. Leibler. «On Information and Sufficiency». In: *Ann. Math. Statist.* 22.1 (1951), pp. 79–86 (cit. on p. 53).

[9] Édgar Roldán and Juan M. R. Parrondo. «Entropy production and Kullback-Leibler divergence between stationary trajectories of discrete systems». In: *Phys. Rev. E* 85 (3 Mar. 2012), p. 031129. DOI: 10.1103/PhysRevE.85.031129. URL: https://link.aps.org/doi/10.1103/PhysRevE.85.031129 (cit. on p. 53).

[10] Miloš Tišma et al. «ParB proteins can bypass DNA-bound roadblocks via dimer-dimer recruitment». In: *Science Advances* 8.26 (2022), eabn3299. DOI: 10.1126/sciadv.abn3299. URL: https://www.science.org/doi/abs/10.1126/sciadv.abn3299 (cit. on p. 53).

[11] Paolo De Los Rios and Alessandro Barducci. «Hsp70 chaperones are non-equilibrium machines that achieve ultra-affinity by energy consumption». In: 3 (May 2014). Ed. by Arup K Chakraborty, e02218. DOI: 10.7554/eLife.02218. URL: https://doi.org/10.7554/eLife.02218 (cit. on p. 53).