PhD Thesis in *Mathematics for Engineering*
Dipartimento di *Scienze di Base ed Applicate per l'Ingegneria*
Facoltà di Ingegneria - Sapienza University, Rome, Italy

(XXX Ciclo - A.A. 2017/2018)

Theoretical models and numerical methods for the study of sub-cellular phenomena

Barbara Coluzzi Bartoccioni

PhD Advisor: Prof. Alberto Maria Bersani

to my mother

This work of thesis would not be possible without the continuous support of my PhD Advisor, Alberto Bersani. I am in particular indebted to him both for introducing me into the research field of Michaelis-Menten kinetics and for helping me into my first attempts of combinatorial calculation. I moreover due the involvement in the idea of studying the possible application of the Renormalization Group for Singularly Perturbed Differential Equations to Michaelis-Menten kinetics to Enrico Bersani, that I gratefully acknowledge for suggestions and support, too. Nonetheless, the results on the numerical study of the Poland-Scheraga model for DNA denaturation transition would not be obtainable without the additional fruitful collaboration with Edouard Yeramian, that started quite a long time ago. Finally, I thank Andrea Giansanti, Alessandro Milanesi, Emanuele Raccah, Gavriel Segre and Pierluigi Vellucci for discussions and encouragements.

Contents

Chapter 1

Introduction

As usual in Biomathematics [1, 2, 3, 4, 5, 6], the present thesis rests on a broad background, that includes numerical simulations, numerical methods and different theoretical approaches. These last ones extend, in the present case, from the Statistical Mechanics [7] and the Renormalization Groups techniques [8, 9] to mathematical methods, such as in particular the standard Perturbation Method for solving ODEs with singular perturbations [10, 2].

This is partially due to the fact that we are concerned with discussing detailed results on two quite different biomathematical problems, *i.e.*, DNA denaturation and enzyme kinetics. Indeed, both of the subjects are essential for a better general understanding of sub-cellular phenomena. Therefore, it appears important to us to dedicate the beginning of this work (Section 1.1) to sketch the picture in which we insert the *fil rouge* that links these two different topics, that we are going to introduce in the following Section 1.2 and Section 1.3, respectively.

1.1 Sub-cellular phenomena: From DNA to enzyme kinetics

When thinking to an organic cell, the biological unit of living organisms, one can be fascinated by very different phenomena, from its capacity to replicate to the one of moving by itself (in some of the cases, such as the one of unicellular organisms). Let us focus on the prototype of cell that we usually encounter in the evolved living animals, that can be, for instance, a cell of: the brain, the heart, the lung, the liver, the dermis (an epidermal layer), the blood, and so on; in other words, a generic specialized cell that usually (apart from the blood red cell, in particular) is provided, first of all, of a nucleus.

The nucleus contains the most of the genetic material. Let us remind that the genes are sequences formed by only four nucleotides: Cytosine (C), Guanine (G), Adenine (A) and Thymine (T). They belong to the two strands, that are arranged in the well known double helix structure, of a long linear DeoxyriboNucleic Acid (DNA) molecule. Furthermore, the DNA molecules are themselves arranged in the chromosomes. Obviously, the chromosomes have a fundamental role both in the single specialized cell replication and in the transmission of the genetic heritage from one individual to its descendants, but these sub-cellular phenomena would bring us too far from the main topics of the present work. Let us instead focus on the classic sub-cellular phenomenon that allows the formation of a specific protein, or of a specific enzyme.

In this process, as in the cell replication, the chromosomes need to unroll, so that the sequence of nucleotides of a single DNA strand can show itself. Nevertheless, for the usual protein / enzyme formation, it is enough the opening of the small (coding) region of the single DNA molecule that contains the relevant gene for the considered protein / enzyme. Indeed, this is already a sub-cellular phenomenon that would require a deep study, since it involves a large number of factors, *i.e.*, the positions of the considered region of the chromosome within the nucleus, the ratio between the concentration of nucleotides G and C to the one of nucleotides A and T in the considered region of

the given DNA molecule, the cell region temperature $\mathcal{T}$ and its hydrogen potential pH, the possible presence of enzymes (apart from the one that will be eventually formed), and so on.

The next step is allowed by the well known basic DNA property, that is shared by the other Nucleic Acid, the RiboNucleic Acid (RNA). In fact, the structure of this last one is due to a sequence formed by four nucleotides, too, three of which are the same as in the DNA, whereas the Thymine is replaced by the Uracil (U). The main difference is that usually the RNA is found in nature as a single strand folded on itself, instead that in the double strand helicoidal structure characteristic of the DNA. In any event, the basic property we were referring to is that the only allowed links are the A-T (in the DNA case) or the A-U (in the RNA one) and the G-C. This implies, first of all, that in particular the DNA molecule can be replicated, within the cell nucleus. Especially, from the present discussion point of view, this implies that the nucleotides of a single strand can be seen, taken three at the time, as the letters of a (redundant) alphabet that, in most of the cases, are in correspondence with a given amino acid, *i.e.*, with one of the basic components of the proteins / enzymes. More correctly, though this subject would need an in-depth study, too, the triplet has to belong to a coding DNA region, and it can also correspond to the "start" or "stop" signal of the coding region (codon).

Then, the formation of the protein / enzyme itself happens through various steps. Very briefly, a kind of RNA, the messenger RNA (mRNA), is involved as the one that reproduces the sequence of DNA triplets, or codons (for instance, the triplet GAC can only transforms itself in CUG), and that carries the information to the ribosome, where the chain of amino acids of the unfolded protein / enzyme is formed. One has at least still to remind that, in fact, the mRNA has itself a double helix structure, and that two other kinds of RNA molecules, the transfer RNA (tRNA) and the ribosomal RNA (rRNA) are involved in the process.

On these bases, one can already see that the topic of DNA denaturation, *i.e.*, the opening in temperature of the DNA coupled strands, is a key point for a better understanding of the DNA opening that one observes in nature, in the context of the sub-cellular phenomena. Indeed, the DNA opening within the cell is the process that allows both DNA replication and the protein / enzyme formation itself. Therefore, the subject is attractive despite the fact that, in the present work, we will face its study only from the point of view of the behaviour of a simple statistical mechanics model, that appears correct for reproducing the denaturation in temperature but disregards other relevant details, such as the double helix structure and the complex mechanism that controls the opening within the cell.

To continue, it would be outside the purposes of the present work to attempt to discuss exhaustively the properties of the enzymes and their different roles within both the intra-cellular and the inter-cellular phenomena. Rather, let us recall the classic situation that one encounters when facing enzyme kinetics, *i.e.*, let us focus on an enzyme, $\mathcal{E}$, that interacts with a substrate, $\mathcal{S}$ (in detail, a molecule or protein), by giving rise to a complex, $\mathcal{C}$. This last one, furthermore, transforms itself in the product, $\mathcal{P}$ (in detail, a different molecule or protein), by releasing, contemporaneously, the original enzyme, too. Thus, the enzyme acts as a catalyst, that accelerates or that simply makes possible the chemical reaction $\mathcal{S} \to \mathcal{P}$. Correspondingly, the complete chemical reaction can be schematised as:

$$\mathcal{E} + \mathcal{S} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathcal{C} \xrightarrow{k_2} \mathcal{E} + \mathcal{P}, \tag{1.1}$$

with associated kinetic constants k_1, k_{-1} and k_2, that can be seen as the rate constants of the reaction.

A very important example, perhaps particularly helpful to easily understand the kind of problem, is the one of the reaction involving as enzyme the hemoglobin. Anyway, it is rather an example of inter-cellular phenomenon that in addition involves an enzyme with more than one (four) active sites. Letting aside that there are many different forms of this enzyme and that it can in principle bind to different gas molecules, usually the hemoglobin, that is found in the red blood cells, binds to oxygen. Thus, in the previous scheme, the oxygen is the substrate and the oxyhemoglobin (hemoglobin saturated with oxygen) the complex. Naturally, the binding happens in the lungs, whereas the

irreversible part of the reaction is the releasing of oxygen to the single cells in the tissues (that also gives back the enzyme in its original state, with unoccupied active sites).

In order to remain at an introductory discussion level, we postpone to the following Chapter 2 and Chapter 3 the detailed presentation of the two main subjects that we study in this work, respectively. Nonetheless, apart from the fact that both of them are relevant to a better general understanding of sub-cellular phenomena, there is another, more subtle, *fil rouge* that links the DNA denaturation to the enzyme kinetics in this thesis. In fact, in both the cases, we will refer in our study to (different) Renormalization Group techniques. Let us, therefore, better introduce the subjects from the theoretical point of view.

1.2 Theoretical approaches

The present study of DNA denaturation transition is based on a simple model that was introduced by Poland and Scheraga [5, 11, 12] (PS) and that dates back to 1966. This PS model appears in fact as a toy model from the Statistical Mechanics point of view [7], particularly in the pure (without disorder) case, in which the link energies of the two possible couples of nucleotides (AT and GC, respectively) are assumed to be the same [13].

Roughly speaking, one can imagine a uni-dimensional Ising chain, in which a given spin $\sigma_i \in \{0, 1\}$ takes the value 1 if the couple of nucleotides (bases) in the corresponding position $i \in \{0, 1, 2, \dots, \mathcal{N}\}$, along the double stranded DNA structure, is linked, hence if the two strands (chains) are linked in that position, whereas it takes the value 0 in the opposite case. In fact, the uni-dimensional Ising model with short range interactions does not undergo any phase transition. Nonetheless, the behaviour of the present PS model displays a phase transition since the entropic effects associated to the open regions (the loops) are taken into account. With regards to these long-range interactions, the model is an example of an "almost" uni-dimensional system. As usual, the presence of a phase transition corresponds to the presence of a singularity in the derivative of the free energy with respect to the temperature, $\mathcal{T}$, at some order, in the thermodynamic limit $\mathcal{N} \to \infty$ (here, $\mathcal{N}$ is to be interpreted as the number of base pairs). For the sake of precision, the transition in the original PS model was predicted to be of second order [5, 11, 12], hence corresponding to a singularity in the second order derivatives of the free energy.

On the other hand, the importance of the effect of self-avoidance (namely, of the fact that different parts of the two DNA chains cannot occupy the same position in the space), in this pure case became clear only more than thirty years later, at the beginning of this century. Indeed, this effect is as much important as to change the order of the transition (*i.e.*, the kind of singularity of the free energy in the thermodynamic limit). In fact, the same PS model undergoes a first order phase transition when the self-avoidance is completely taken into account. Interestingly, this was theoretically predicted [14, 15] within a Renormalization Group approach, by making use of theoretical results on polymer networks that had been obtained meanwhile [16, 17].

Moreover, on the general bases of various other Renormalization Group predictions (see in particular [18, 19]), one could expect that the effect of disorder were to smooth the transition in this case. Nonetheless, this was difficult to be proved, since the PS model with self-avoidance is a quite peculiar polymer model. Importantly, mathematical results [20, 21], that are obtained within a probabilistic approach and that date back to 2006, bring to the conclusion that the disordered (d) model undergoes a smoother transition than the pure (p) one. Anyway, the detailed expected behaviour of such a kind of disordered models is still to be better theoretically understood [22, 23, 24]. In particular, to our knowledge, there are no theoretical predictions on the value of the smooth transition critical exponent.

On the other hand, when studying the experimental results on DNA denaturation [25], that mainly regard experiments performed on a single sequence of base pairs of a given length $\mathcal{N}$, one observes that the measured density of closed base pairs, $\Theta(\mathcal{T})$, has a multi-step behaviour. In other words,

$\Theta(\mathcal{T})$ displays abrupt changes as a function of $\mathcal{T}$, though these are not mathematical discontinuities, since they are in any event smoothed out by the finite chain length. Correspondingly, sharp peaks in the derivative of $\Theta(\mathcal{T})$ with respect to $\mathcal{T}$, that has the same behaviour as the specific heat, are present in these points. These experimental results suggest a "sharp" transition, that was ordinarily interpreted to correspond to a first order one, from the Statistical Mechanics point of view.

Therefore, in addition to some conflicting results of numerical simulations that we will discuss in Chapter 2, there is an apparent contradiction, since the presence of a first order transition in the PS model with self-avoidance, also in the disordered case, could appear necessary for the model being able to correctly capture the experimental observations. Our present analysis, apart from confirming the disorder relevance in the thermodynamic limit, in agreement with the recalled mathematical results, is to be interpreted just as giving the possibly correct answer to this apparent contradiction. At the same time, this answer would solve the problem that is posed by the previous conflicting results of the numerical simulations, too.

Without entering more in detail into this subject for the moment, it is to be underlined that we encounter Renormalization Group techniques, within the DNA denaturation context, as successful approaches both for explaining the first order character of the transition in the case without disorder and for predicting the disorder relevance.

Taking into account such a noticeable effectivity of the Renomalization Group techniques, we find particularly interesting, in the second part of this work, after devoting the introductive Section 3.1 to recall these techniques in generic terms, to test a Renormalization Group approach in the different context of dynamical processes. The approach was proposed by Chen, Goldenfeld and Oono [26, 27], for the study of Singularly Perturbed Differential Equations (SPDERG). In detail, we consider the case of boundary layer problems, *i.e.* the ones that are encountered in enzyme kinetics and that are usually dealed with by the standard Perturbation Method for singular ODEs. Importantly, though we make partially use of numerical methods, we test the approach in a rigorous way, that allows to underline similarities and differences between the SPDERG and the standard Perturbation Method from the mathematical point of view.

A milestone work in the mathematical modelling of enzyme kinetics is the one by Michaelis and Menten [28, 29], that introduced, more than one century ago, the key ingredients of the standard Quasi Steady State Approximation (sQSSA) on the basis of accurate experimental results. This is a possible starting approximation to the system of ODEs that describes the chemical reaction (1.1). Indeed, it is the approximation that is still the most largely used for fitting experimental data at large times [30].

One generally refers to (1.1) and to the system of ODEs that describes it, as to a simple example of Michaelis-Menten (MM) enzyme kinetics. In order to introduce the subject of our study in this context, we recall that such a system of ODEs can be reduced, because of conservation laws, to two first order ODEs with given initial conditions. As we will discuss in detail in Chapter 3, there are two main Quasi Steady State Approximations [29] that differ in the choice of the independent variables. In order to test the SPDERG approach, we apply it within the sQSSA framework. Our motivation is that the other main approximation, *i.e.*, the total Quasi-Steady State Approximation (tQSSA), though more useful from the experimental point of view, is more involved from the analytical one.

In the sQSSA, one takes as independent variables the Substrate $\mathcal{S}$ and the Complex $\mathcal{C}$ and one appropriately rescales the variables in the equation, the time t being included, to the aim of making dimensionless the ODE system. Though there is some freedom of choice, this last passage leads to the introduction of a dimensionless ε parameter, that is ordinarily taken to be equal to the ratio between the initial Enzyme concentration, e_0, and the initial Substrate one, s_0, *i.e.*, $\varepsilon = e_0/s_0$. This parameter, in usual *in vitro* experiments, is taken much less than 1.

Roughly summarizing, at the end of this procedure, the ODE system turns out to consist of a singularly perturbed ODE. This is the ordinary situation in which one applies the standard Perturbation Method [10, 1, 29]. In fact, this ODE system has to be regarded as the outer system that

describes the original dynamical process at large times and that allows to obtain the outer solutions. In order to study the short initial time region, one has furthermore to make the transformation $t \to \tau = t/\varepsilon$ and to consider the corresponding inner ODE system, too, that is regularly perturbed and that allows to obtain the inner solutions.

On the basis of the standard Perturbation Method, both the inner and the outer solutions can be written perturbatively, as expansions in the parameter ε. Therefore, at each order, one calculates the corresponding terms in the expansions by solving the inner and outer ODE system at that order. Finally, the best analytical approximations to the correct solutions that are obtainable from the results, and that are known as uniform approximations, are given by the sum of the inner and of the outer solutions up to the considered order in ε, minus the common terms.

This brief discussion of a subject that we will recall item by item, in particular in the case of MM enzyme kinetics beyond the sQSSA [2, 31], has the aim of clarifying from the beginning the main differences between the standard Perturbation Method and the SPDERG approach. In fact, in this last case, one restricts the study to the inner ODE system. As we will explain in Section 3.4, the approach is based on the renormalization of appropriate integration constants, that generally need to contain the bare initial conditions, though in our rigorous *ad hoc* application to MM kinetics, to which, more correctly, we refer with the shortening SPDERG, we directly renormalize the initial conditions themselves. By this way, one usually obtains the leading terms of the outer solutions up to the considered order in ε, too [26, 27]. Moreover, one directly gets the SPDERG uniform approximations to the correct solutions.

1.3 An anticipation of our main results

- The contribution to a better understanding of DNA denaturation transition that is discussed in this work consists of two parts:
  - i) A conclusive analysis of numerical simulation data on a simple disordered model *à la* Poland-Scheraga, whose results are presented in [32], a work in collaboration with E. Yeramian. In this study, first of all, we confirm the numerical evidence for a smoother transition in the thermodynamic limit in the disordered case and we give a refined evaluation of the critical exponent.
  - ii) A combinatorial approach, in which we rebuild the calculation of the probability of finding a given subsequence of length L in a chain of length $\mathcal{N}$, $P(L,\mathcal{N})$, in the case of an uniform binomial distribution for the two possible base pairs. As suggested in [32] and as we will explain more in detail, this quantity appears to have a key role for understanding "up to which point" one can expect to find evidence for disorder relevance, at fixed model parameters, depending on the finite chain length $\mathcal{N}$ that one is looking at.

  To the aim of sketching the main results in the first item, it is worth noting that the present numerical study was possible because of the approximation, that can be made as correct as required up to very large l values, of a power law of the kind $1/l^{c_p}$ with a sum of exponentials. This SIMulations with EXponentials (SIMEX) scheme is originally based on in-depth mathematical analyses [33, 34]. Despite the fact that the implementation, in the present relatively simple case of a linear DNA molecule model *à la* PS, turns out to give an algorithm that is basically the same as another one already present in the literature [35], the SIMEX scheme is indeed more generally applicable.

  In the present work [32] and in previous results on the same subject [36, 37], in which a standard data analysis was considered, we performed numerical simulations and we applied finite size scaling techniques in order to infer the thermodynamic limit model behaviour from the data. It is the first time, to our knowledge, that the finite size scaling approach, that is based on

Renormalization Group results [8], is considered in the case of a disordered model for DNA denaturation, in particular by using the SIMEX algorithm [37, 32].

As we will discuss in detail, despite mathematical results for a class of models that should encompass the present one [20, 21], the possibility of a peculiar model behaviour in the thermodynamic limit is suggested by other numerical results on a slightly different realization of the same model [38, 39]. Moreover, this possibility was not ruled out from our previous standard finite size scaling analysis [37], since this possible behaviour could be made evident only by an *ad hoc* finite size scaling analysis.

Therefore, our results can be interesting from the point of view of definitely confirming that also a refined, *ad hoc*, finite size scaling analysis of the data is in agreement with disorder relevance in the present case. Importantly, the disorder relevance has to be interpreted as usual, *i.e.*, in the thermodynamic limit the disordered model undergoes a smoother transition than the pure one. Moreover, the present kind of analysis allows a refined evaluation of the critical exponent.

Finally, in a semi-quantitative framework, we suggest an explication that would solve the paradox due to the conflicting previous numerical findings, and that could allow to make useful predictions from the experimental point of view, since they would regard the finite chain length behaviour of the model. This is the subject to which we start to dedicate our attention subsequently.

- On the other hand, the contribution to a better understanding of MM enzyme kinetics that is presented in this work consists of two parts, too, that roughly correspond to the two different Quasi Steady State Approximations that one can take as starting point:

  - i) We propose an *ad hoc* rigorous way of applying the SPDERG approach to the particularly demanding case of MM enzyme kinetics and we make an in-depth analysis of similarities and differences with the standard Perturbation Method, in the framework of the sQSSA. The results of this study are presented in [40], a work in collaboration with A.M. Bersani and E. Bersani.
  - ii) We briefly discuss the different case of the tQSSA framework, that is more interesting from the experimental point of view: we recall the ODE system that is the starting point in this framework and we outline both the possible simplifications and the possible difficulties in applying the SPDERG approach in such a context. In fact, this is our next work project.

  In order to deepen the main results in the first part, we need to anticipate that in the original works [26, 27], in which the SPDERG approach is proposed and tested, they were considered only cases of second order ODEs where one could intuitively understand the appropriate constants to be renormalized. In the present case of MM enzyme kinetics, one can naturally write the considered system, that consists of two first order ODEs, as a single second order ODE, too. Nevertheless, it does not appear that such a second order ODE could be dealed with, within the SPDERG approach, by choosing appropriate integration constants. From this point of view, MM enzyme kinetics appears a particularly demanding situation for applying the approach, as it is in particular made evident by a previous attempt in this direction within the tQSSA framework [41].

  Correspondingly, our first main result is to find out an *ad hoc*, rigorous, way of applying the SPDERG approach in the present case, by starting from the less analytically demanding sQSSA framework. Indeed, our method simply consists in starting from the two first order ODE system and, as advanced, in directly renormalizing the initial conditions. This allowed us to make an in-depth analysis of similarities and differences with the standard Perturbation Method. The analysis is rigorous from the mathematical point of view, and its conclusions can be taken for

granted in future similar studies. In particular, this work makes possible the application of the SPDERG within the tQSSA framework, that we introduce here.

It appears also important to underline from the beginning that we are able to apply the SPDERG approach starting from the sQSSA by explicitly determining the second order terms in the expansion parameter ε. Therefore, in particular the inner solutions are computed up to the second order for the first time to our knowledge. Furthermore, we go beyond another obstacle in the SPDERG application to MM enzyme kinetics, the one of a possible non zero limit of the solutions at large times. In fact, we propose appropriately refined SPDERG uniform approximations, that automatically approach zero at large time, on the basis that some characteristics of the solutions that emerge from our study could be iterated. Summarizing, we propose and test an *ad hoc* way of applying this Renormalization Group technique that turns out to be particularly appropriate to deal with MM kinetics. We refer with the shortening SPDERG to this *ad hoc* method of applying such an approach.

The thesis is divided in the following way: after the present first introductive Chapter, in the second Chapter: in Sections 2.1-2.6 we recall the known results on the PS model for DNA denaturation transition, both in the pure and in the disordered case; In Section 2.7, after recalling the used SIMEX algorithm (Subsection 2.7.1) we present the original results of this thesis work, *i.e.*, the finite size scaling analysis of the numerical data in terms of pseudo-critical temperatures (Subsections 2.7.2-2.7.4) and the careful analysis of the data on the loop length probability distribution at different temperatures (Subsection 2.7.5), both of which suggested the introduction of the crossover length scale $\mathcal{N}^*$ and of an appropriate phenomenological scenario in order of evaluating it (Subsection 2.7.6); in Section 2.8 we present our original combinatorial calculation of $P(L, \mathcal{N})$ and the preliminary results on the refined evaluation of $\mathcal{N}^*$ that can be obtained by the knowledge of this quantity. In the third Chapter: in Section 3.1 we recall a few basic Renormalization Group ideas; in Sections 3.2-3.3 we recall the known Perturbation Method results within the sQSSA framework; in Section 3.4 we introduce the SPDERG approach; in Sections 3.5 we discuss the original results of this thesis work, *i.e.*, the first order SPDERG contribution (Subsection 3.5.1), the second order one (Subsection 3.5.2), our refined SPDERG second order approximations (Subsection 3.5.3) and the comparison between the different best approximations (Subsection 3.5.4); in Section 3.6 we discuss in some detail the different tQSSA framework and we outline both the possible simplifications and the possible difficulties in applying the SPDERG approach. In the fourth Chapter, we present our Conclusions. The work is completed by two Appendices.

Chapter 2

The Poland-Scheraga model for DNA denaturation transition

The DNA denaturation transition is a topic with a long history in the scientific literature and its correct modelling is still a debated subject [42]. Here, before presenting our work, we limit ourselves to recall the key results on simple models *à la* Poland-Scheraga (PS) [5, 11, 12] with self-avoidance, by starting from the pure case and by discussing in some detail the possibly different theoretical predictions on the effect of disorder, in order to sketch the picture of the correct background that motivated our analyses.

2.1 The pure PS models: A statistical mechanics point of view

The first attempts to model DNA denaturation, *i.e.*, the helix coil transition, led to a uni-dimensional Ising model with short range interactions [43], that cannot display any true transition in the thermodynamic limit. This situation changes when taking into account the entropic effects that are due to the loops (the coiled regions), which led to the PS model [5, 11, 12]. This "almost" uni-dimensional model, that is an off-lattice DNA model, in the initially considered case in which the entropy of the loop is that of a $3D$ (we label D the spatial dimension) Random Walk (RW) of the same length, displays a second order transition, as extensively explained in [13], within a more general picture.

In order to fix the ideas, let us present, in [Fig. 2.1], a realization of the PS model, just in the original Random Walk case (RWDNA), on a $2D$ square lattice, *i.e.*, an example of configurations of an on-lattice DNA model *à la* PS. In detail, let us think to a couple of RWs of the same length $\mathcal{N}$, that live on a $2D$ lattice and that share the same origin, in which the couple of monomers (bases) in the same position i ($i \in [0, 1, 2, \ldots, \mathcal{N}]$) along the chain gains an energy ϵ *only if* they occupy the same position on the lattice, too. Notice that the configuration in the figure is chosen in such a way to be an *ad hoc* example. Thus, it is possibly peculiar with respect to the allowed ones. The model is in particular supposed to be near the transition point, and the length of the free ends is taken to be zero.

With the aim of writing down the canonical partition function, $Z_{\mathcal{N}}^{RWDNA}$, let us consider the single model configuration as a succession of k helix regions of length n_h and coil regions of length $2l_h$, plus the free ends (RWFE). Then, let us label with: i) $\eta_{n_h}^{RW}$ the number of RW of length n_h; ii) $\eta_{2l_h}^{RWL}$ the number of RW loops (RWL) of length $2l_h$; iii) η_{2r}^{RWFE} the number of RW of length $2r = \mathcal{N} - 2\sum_{h=0}^{k}(n_h + l_h)$ (*i.e.*, the free ends length). Then, $Z_{\mathcal{N}}^{RWDNA}$ can be written as:

$$Z_{\mathcal{N}}^{RWDNA} = \sum_{\{n,l\}} e^{\beta\epsilon}\eta_{2r}^{RWFE}\prod_{h=0}^{k} e^{\beta\epsilon n_h}\eta_{n_h}^{RW}\eta_{2l_h}^{RWL}. \tag{2.1}$$

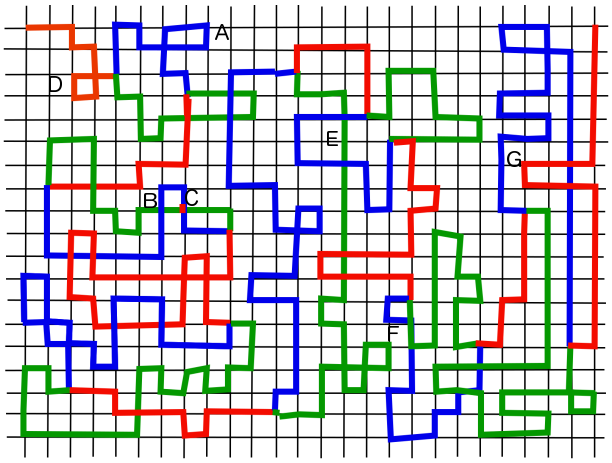

Figure 2.1: **A configuration of a RWDNA, a model *à la* PS, on a $2D$ square lattice.** The helix regions are colored in red, whereas the two strands of the coil regions are (arbitrarily) colored in green and blue, respectively. The model allows monomers to occupy the same lattice point, for instance, also if: i) they belong to the same strand within a coil region (as in A); ii) they belong to the two different strands, within the same coil region (as in B); iii) they belong to the same helix region (as in D); iv) a couple of them belongs to a helix region and a third monomer to a coil one (as in G); v) they belong either to different coil regions (as in E), or to different helix regions (not shown). Moreover, the different parts of the two chains could, for instance, overlap on lattice points that are nearest neighbours (as in F). On the other hand, the configuration gains an energy ϵ only if the two monomers that occupy the same lattice point have the same position along the two strands (as in all the helix regions and, in particular, in C). See the text for details.

Here, $\beta = 1/\mathcal{T}$ is the inverse temperature (by taking a Boltzmann constant $K_B = 1$ for simplicity) and the sum runs over all possible partitions into k helix region (segments) and k coil ones (loops). For the sake of clarity, at the risk of being repetitive, we outline that we are implicitly taking the total segment length equal to $n = \sum_{h=0}^{k} n_h$ and the total loop length equal to $2l = 2\sum_{h=0}^{k} l_h$, with $0 \leq 2(n+l) \leq \mathcal{N}$, whereas the number of helix / coil regions in the succession, k, depends on the considered partition in the $\{n, l\}$ set. Moreover, also that the two chains share the same origin is taken into account by the factor $e^{\beta\epsilon}$. Finally, such a way of writing $Z_{\mathcal{N}}$ is suitable for understanding its generalization to the self-avoiding walk (SAW) case.

The model is solvable in the grand canonical ensemble. In fact, by introducing the fugacity z, one can calculate the grand canonical partition function, $G^{RWDNA}(z)$, since one can invert the order of the sum and of the product, by obtaining that the succession of alternating segments / loops gives rise to a geometric series. One finds:

$$G^{RWDNA}(z) = \sum_{\mathcal{N}=0}^{\infty} z^{\mathcal{N}} Z_{\mathcal{N}}^{RWDNA} = \frac{G^{RWFE}(z) G^{RWS}(z, \beta\epsilon)}{1 - e^{\beta\epsilon} G^{RWL}(z) G^{RWS}(z, \beta\epsilon)}. \tag{2.2}$$

In this last formula, we are labeling by $G^{RWS}(z, \beta\epsilon)$, $G^{RWL}(z)$ and $G^{RWFE}(z)$, the grand canonical partition functions of the segment, of the loop and of the free ends, respectively.

Without entering into the details of the analytical calculations [13], let us recall that the behaviour, in the thermodynamic limit $\mathcal{N} \to \infty$, is determined by the singularity in z that is closest to the origin. In the present case, the model undergoes a transition since there is a value, greater than zero, of the temperature (here, of the scaled variable $\beta\epsilon$), at which the singularity that is closest to the origin switches from the one of $G^{RWFE}(z)$, in the numerator of $G^{RWDNA}(z)$, to the one of the denominator. Correspondingly, the high temperature behaviour is that of two free RWs with the same origin, whereas the low temperature one is that of the coupled chains, or, in other words, of a single segment.

For the sake of correctness, in the original PS model, the two chains are always linked both in the origin and in the end points, as in the configuration of the on-lattice DNA model *à la* PS that is

shown in [Fig. 2.1]. Therefore, in the high temperature / low temperature phases the model behaves as a single open loop / a single segment, respectively. Nevertheless, from the analytical point of view, the free ends are to be considered as a boundary condition, that does not influence the critical behaviour.

It is to be noted, moreover, that in the on-lattice DNA models the critical point is in fact a tri-critical point in the $(z, \mathcal{T})$ plane [44], characterized by the tri-critical exponent ϕ. However, let us rather recall the main analytical results on the original PS model, where one disregards the lattice structure itself, by taking the geometrical dimension $\mathcal{D}$ that the on-lattice polymer "occupies" equal to 1. In such a case, the whole model thermodynamic behaviour can be related to the one of the probability distribution $P(\mathcal{T}, l)$ of the loop length $2l$ and to the value of its exponent c_p (the label p underlines that we are in the pure case):

$$P(\mathcal{T}, l) \propto \frac{e^{-l/\xi(\mathcal{T})}}{l^{c_p}}. \tag{2.3}$$

Here, $\xi(\mathcal{T})$ is the thermal correlation length.

When approaching the critical point, $\mathcal{T}_c$, from the low temperature phase, the thermal correlation length diverges as $\xi(\mathcal{T}) \sim (\mathcal{T}_c - T)^{-\nu_p}$ and, correspondingly, the order parameter, that is the density of closed base pairs, $\Theta(\mathcal{T})$, approaches zero as $\Theta(\mathcal{T}) \sim (T - T_c)^{\beta_p}$, with $\beta_p = \nu_p - 1$. In fact, the only independent exponent is c_p, and one has $\nu_p = 1$ (hence $\beta_p = 0$) for $c_p \geq 2$, whereas $\nu_p = 1/(c_p - 1)$ for $c_p < 2$ [13]. Therefore, one would observe a first order transition for $c_p \geq 2$, with the order parameter that undergoes a discontinuity at $\mathcal{T}_c$, in the thermodynamic limit $\mathcal{N} \to \infty$. On the other hand, in the original PS model, the value that is assumed for c_p is inferred from the known behaviour of a $3D$ RW loop, *i.e.*, $c_p = 1.5$, that gives, as already recalled, a second order transition [11, 12].

To better sketch the relation between the off-lattice PS models and its possible realizations on a D-dimensional lattice, one has to remind that the considered polymer geometrical correlation length, $\xi_{\mathcal{G}}$, is to be taken into account in the last case, too, with $\nu_{\mathcal{G}} = 1/\mathcal{D}$ its critical exponent. One can show [44] that $\phi = \nu_{\mathcal{G}}/\nu_p$ whereas the hyper-scaling relation for the specific heat critical exponent, α_p, is $\alpha_p = 2 - \mathcal{D}\nu_p = 1 - 1/\phi$. Hence, in the tri-critical first order case of $\phi = 1$, the thermal correlation length exponent, ν_p, is equal to the geometrical correlation length one, $\nu_{\mathcal{G}}$, [44] (coherently, this corresponds again to $c_p \geq 2$). On the other hand, in the simpler off-lattice models, one has directly $\alpha_p = 2 - \nu_p$ [13].

The aim of the previous discussion was to make evident the importance of the c_p value. From the experimental point of view, when studying the denaturation of DNA molecules, one observes abrupt changes in the density of closed base pairs as a function of $\mathcal{T}$ (the multi-step behaviour). Correspondingly, the derivative with respect to the temperature of this quantity, that behaves as the specific heat, displays sharp peaks. Though a real DNA sequence is first of all a *disordered* one (since the link energies are largely different between the AT and the GC couples of base pairs), it is for these reasons that one expects that a correct statistical mechanics model should undergo a first order transition in the thermodynamic limit, at least in the pure case.

From formula (2.3), one has that at the critical point, where the thermal correlation length diverges, $P(\mathcal{T}_c, l) \sim 1/l^{c_p}$. Hence, c_p measures the power law decreasing probability of observing large loops. Nonetheless, these loops model the ones formed in the space by the two interacting chains, that are both self-avoiding and mutually avoiding. From this point of view, to determine the c_p value on the basis of the behaviour of a single free RW loop is an approximation. In fact, in the same period in which the PS model was introduced in [11, 12], it was observed by Fisher [45] that one can more correctly choose the exponent on the basis of the behaviour of a single free Self-Avoiding Walk (SAW) loop, by getting $c_p \simeq 1.74$. Moreover, it was remarked in the same work that the expected transition would be still sharper if the self-avoidance effects were more completely taken into account.

With the aim of clarifying this last point, that is particularly important in our work, we present

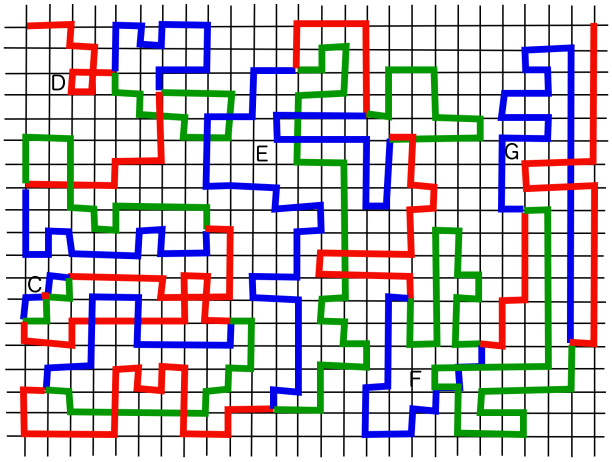

Figure 2.2: **A configuration of a RWDNA with SAW loops, a model *à la* PS with self-avoidance partially taken into account, on a $2D$ square lattice.** As in [Fig. 2.1], the helix regions are colored in red, whereas the two strands of the coil regions are (arbitrarily) colored in green and blue, respectively. The model does not allow monomers to occupy the same lattice point if: i) they belong to the same strand within a coil region (as in A, in [Fig. 2.1]); ii) they belong to the two different strands, within the same coil region (as in B, in the previous [Fig. 2.1]). Nonetheless, the model does still allow monomers to occupy the same lattice point, for instance, if: i) they belong to the same helix region (as in D); ii) a couple of them belongs to a helix region and a third monomer to a coil one (as in G); iii) they belong either to different coil regions (as in E), or to different helix regions (not shown). Moreover, the different parts of the two chains could, for instance, overlap on lattice points that are nearest neighbours (as in F). On the other hand, the configuration gains an energy ϵ only if the two monomers that occupy the same lattice point have the same position along the two strands (as in all the helix regions and, in particular, in C). See the text for details.

in [Fig. 2.2] a configuration of the RWDNA model *à la* PS with SAW loops. The configuration appears on purpose similar to the one shown in [Fig. 2.1]. Nevertheless, here, the possibility for two monomers to occupy the same lattice point if they belong either at the same or at the two different strands in a given loop is forbidden. Therefore, there are no points as, for instance, the ones that we labeled with A or B in the previous [Fig. 2.1]. Consequently, the total number of configurations, for a given chain length $\mathcal{N}$, is lower. In fact, in these figures, we are presenting possibly peculiar configurations, on a small finite volume lattice. Moreover, we did not impose to the couple of chains in the different figures to have equal total length $\mathcal{N}$. Instead, we paid specific attention to the basic rule, that the two chains can form a segment (the configuration gains the energy ϵn_h, with n_h the segment length) *only if* the monomers that occupy the same points on the lattice have the same positions in the respective chains, too. This is equivalent to require that the portions of the two chains that contribute to a given loop have the same length. To further explain the figures, we underline that the segment can *a priori* have length $n_h = 1$, as, for instance, in the point that we labeled by C.

The figures and this discussion of their details should make more intuitively understandable that, from the analytical point of view, the difference between the RWDNA with RW or SAW loops is contained in the higher value of the exponent c_p in the last case, *i.e.*, the probability of large loops becomes smaller, as can be seen from (2.3). In fact, when looking at (2.1) and (2.2), it should be evident that the RWDNA with SAW loops can be solved within the grand canonical ensemble, too, where once again the succession of segments and loops gives rise to a geometric series, without approximations.

Importantly, if we were instead to take self-avoidance completely into account, the situation would be different. We present in [Fig. 2.3] a configuration of a possible realization of this SAWDNA model *à la* PS, on a $2D$ square lattice (the model was originally proposed and numerically studied in $3D$ in [44]). Also this *ad hoc* configuration appears on purpose similar to the ones shown in [Fig. 2.1] and

[Fig. 2.2]. On the other hand, here the "almost" uni-dimensional structure of the models *à la* PS is the most evident. In fact, one can easily recognize the succession of segments and loops (we are again disregarding the free ends for simplicity), that do never intersect the ones with the others. In other words: i) each single loop is self-avoidant, as already in the previous [Fig. 2.2]; ii) the configuration cannot contain points such as, for instance, the ones that we labeled with D, E, F and G in the previous [Fig. 2.1] and [Fig. 2.2]. This can be more correctly summarised by the simple basic rule of the SAWDNA: two monomers can occupy the same lattice point *if and only if* they are in the same position in the two chains (and, in this case, the configuration gains an energy ϵ) [44].

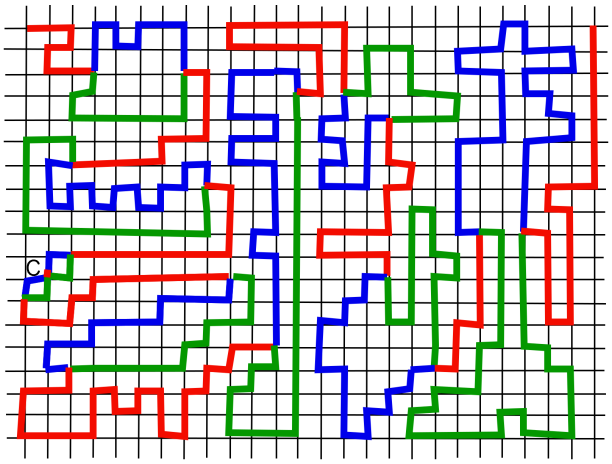

Figure 2.3: **A configuration of a SAWDNA, a model *à la* PS with self-avoidance completely taken into account, on a $2D$ square lattice.** As in [Fig. 2.1] and [Fig. 2.2], the helix regions are colored in red, whereas the two strands of the coil regions are (arbitrarily) colored in green and blue, respectively. The model does not allow monomers to occupy the same lattice point if: i) they belong to the same strand within a coil region (as in A, in the previous [Fig. 2.1]); ii) they belong to the two different strands, within the same coil region (as in B, in the previous [Fig. 2.1]; iii) they belong to the same helix region (as in D in the previous [Fig. 2.2]); iv) a couple of them belongs to a helix region and a third monomer to a coil one (as in G in the previous [Fig. 2.2]); v) they belong either to different coil regions (as in E in the previous [Fig. 2.2]), or to different helix region (not shown). Therefore, the different parts of the two chains cannot overlap on lattice points that are nearest neighbour the one of the other (as in F in [Fig. 2.2]), too. More exactly, two monomers can occupy the same lattice point *if and only if* they have the same position along the two strands (as in all the helix regions and, in particular, in C). Once again, in this last case, the configuration gains an energy ϵ. See the text for details.

To make a few last comments on the figures, we outline that we have explicitly taken into account all the most likely overlaps: of a single loop with itself; of two different loops the one with the other; of a loop with a segment; and of a segment with itself or with another segment. First of all, for the sake of correctness, we underline that the self-avoidance of the segment with themselves can be taken into account analytically, *i.e.*, one can consider in (2.1) and (2.2) the partition function of self-avoiding segments. This corresponds to introduce a further exponent that, anyway, does not influence the critical properties of the model [13]. Moreover, one has to note that already the intersection of a loop with a segment can be expected to happen with lower probability than the one of a loop with itself or with another loop, since it involves three monomers (though two of them are assumed to be coupled, that is an energetically favoured configuration at low temperatures). On the basis of this kind of reasoning, one can expect practically negligible probabilities for intersections involving large number of monomers. In any event, none of these overlaps is allowed in the on-lattice SAWDNA model. In addition, in more realistic similar models, one should take into account first of all the different stiffness of the segments with respect to the loops. Indeed, the stiffness effect on the SAWDNA has been numerically studied, too, and it has been found not to affect the thermodynamic behaviour [46].

When solving the PS model in the grand canonical ensemble, one is neglecting precisely the

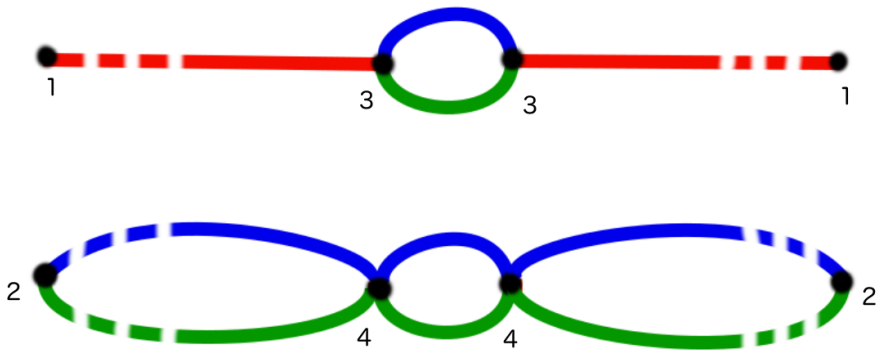

Figure 2.4: **The two SAW networks that allow the theoretical estimate of the effective exponent c_p.** Here the figure follows the ones in [15]. On the top, the SAW loop is embedded between two, definitely larger, SAW chains. On the bottom, the same SAW loop is embedded between two other, definitely larger, SAW loops. The black dots correspond to the vertices and the numbers correspond to their degrees. Within each of the two SAW networks, one can relate the effective exponent c_p to the vertex scaling dimensions (that depend in particular on the vertex degree). In $3D$, one finds: $c_p \simeq 2.115$ from the top scheme; and $c_p \simeq 2.22$ from the bottom scheme [15].

interactions between the different components of the partially coupled chains, *i.e.*, in detail, between different loops, between different segments, and between loops and segments. Therefore, the correct way to take the self-avoidance completely into account in the analytical calculations does not appear evident.

The difficulty can be overtaken [14, 15] by using the key role of the loop probability distribution (2.3) exponent c_p, whose knowledge allows to predict the whole thermodynamic behaviour of models *à la* PS generally. In fact, the correct c_p value to be chosen in the SAWDNA case is an effective exponent that takes also into account the self-avoiding interaction of the SAW loop with the rest of the double chain. The theoretical estimate was obtained within a Renormalization Group approach [14, 15], by using previous analytical results on polymer networks [16, 17]. Roughly summarizing, one can schematise the problem as the one of a SAW loop of length l embedded between two self-avoiding chains of length L, in the case $l \ll L$, and one can get the effective exponent c_p from the known ones that characterize the vertices of this kind of polymer network in such a limit. In order to understand up to which point the approximation is correct, one can moreover consider the opposite case of a SAW loop of length l embedded between two other SAW loops of length L, once again for $l \ll L$. We sketch both of the situations in [Fig. 2.4].

In $3D$, where the vertex exponents are known by an ε-expansion from $D = 4$, these theoretical estimates predict an effective exponent $c_p \in [2.115, 2.22]$. When taking into account that the first of the two considered polymer networks is more realistic than the second one, these estimates are in perfect agreement with the evaluation obtained by numerical simulations in [46], $c_p \simeq 2.15$. In any event, these c_p values are definitely larger than 2. On the basis of the previous discussion, this implies that the pure (homogeneous) $3D$ SAWDNA undergoes a first order transition, as numerically observed in [44], in particular from the scaling behaviour of the order parameter and of the maximum of the specific heat. In fact, the SAWDNA undergoes a first order transition in $2D$, too, where c_p can be theoretically estimated within the same approach and the vertex exponents of the polymer networks are known from an application of Conformal Field Theory.

We resume in [Fig. 2.5] the $3D$ qualitative behaviour of the main observables of the SAWDNA model, that is known on the bases both of the theoretical results in [14, 15] and of the numerical ones in [44, 46]. In detail, we plot: i) the density of closed base pairs as a function of the temperature,

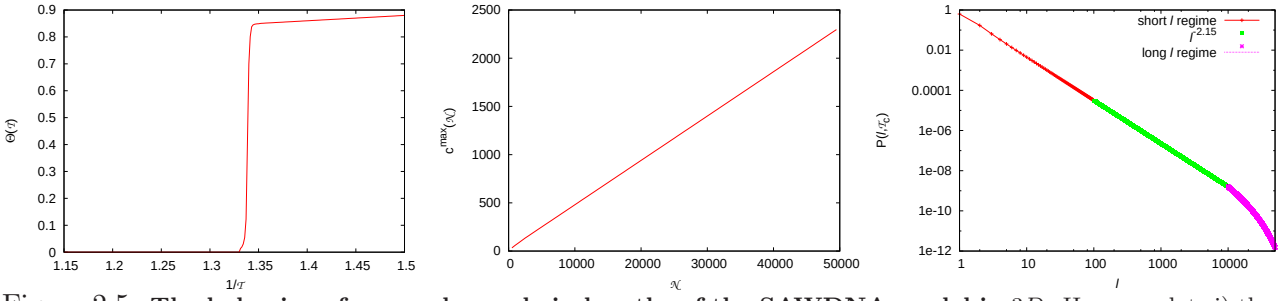

Figure 2.5: **The behaviour for very large chain lengths of the SAWDNA model in** $3D$**.** Here we plot: i) the density of closed base pairs as a function of the temperature, $\Theta(\mathcal{T})$, that displays an abrupt change at $\mathcal{T} = \mathcal{T}_c$ (though it approaches the value $\Theta(\mathcal{T}) = 1$ only at very low temperatures); ii) the maximum of the specific heat as a function of the chain length, $c^{max}(\mathcal{N})$, that is linearly diverging; iii) the (normalized) probability distribution of the loop length, $P(l, \mathcal{T})$, at the critical point $\mathcal{T} \simeq \mathcal{T}_c$, in logarithmic scale, that has slope equal to $-2.15l$ on a large l-range. See the text for details.

$\Theta(\mathcal{T})$; ii) the maximum of the specific heat as a function of the chain length, $c^{max}(\mathcal{N})$; iii) the probability distribution of the loop length, $P(l, \mathcal{T})$, at the critical point, $\mathcal{T} \simeq \mathcal{T}_c$. For the sake of simplicity, here we are ideally observing the very large length behaviours (by taking moreover $\epsilon = 1$). In agreement with a first order phase transition in the thermodynamic limit and with the known c_p value: i) the order parameter, $\Theta(\mathcal{T})$, displays an abrupt change at $\mathcal{T}_c$; ii) the maximum of the specific heat, $c^{max}(\mathcal{N})$, is linearly diverging with the chain length; and, moreover, $P(l, \mathcal{T}_c) \propto l^{-2.15}$ on a large l-range.

2.2 A disordered model *à la* PS

In the models *à la* PS is completely neglected, for instance, the well known double helix structure that characterizes DNA molecules. Therefore, such models are unable, by definition, to reproduce the DNA denaturation due to torsion. Nonetheless, in the models that we discussed in the previous Section 2.1, the most evident experimental missing characteristic is the presence of disorder.

Naturally, one can think to homogeneous synthetic sequences, *i.e.*, to sequences of $\mathcal{N}$ base pairs of the same kind (all AT base pairs or all GC base pairs). The previously discussed SAWDNA model is expected to be particularly suitable for qualitatively describing the denaturation transition in such cases. We outline that the probability distribution of the loop length does not appear to be an observable easy to be experimentally measured. On the other hand, to our knowledge, unfortunately, experimental data on the behaviours, for homogeneous synthetic sequences of different lengths, of the order parameter and of the specific heat maximum are still lacking, too, whereas they would allow to confirm the first order character of the transition in the pure case.

In any event, the structure of an experimental DNA molecule is substantially related to a given sequence of AT and GC base pairs, that have largely different linking energies. In fact, there is an extended literature [5, 25, 42] that is mainly dedicated to the rigorous comparison, for single sequences, of the quantitative behaviours of the main experimental observables with the ones obtained from the PS model. In these works, one usually takes the value of the exponent c_p that is inferred from the free SAW loop, $c_p \simeq 1.74$. In detail, we refer the reader to [42] for a recent discussion on the effect due to the use of the different value $c_p \simeq 2.15$ in a given case, and to [47] for a more general discussion of the differences between the use of this value for the exponent and of the standard one. However, we recall that, in order to correctly reproduce the experimental behaviour, one has to appropriately choose the other parameters of the PS model. For instance, in such a kind of comparisons, one introduces linking energies that depend not only on the considered base pair but also on its first neighbours in the sequence. Moreover, one makes use of a cooperativity factor, σ_0, in order to entropically penalize the opening and closing of the loops.

Indeed, there is a very good agreement between the PS model behaviour and the experimental single sequence data in most of the studied cases [25]. Nonetheless, the introduction of a large number of parameters, apart from the basic exponent that rules the model thermodynamic behaviour, makes difficult to understand the theoretical meaning of such an agreement.

Therefore, here we follow a different line of reasoning, since we are interested, for the moment, in the behaviour of the disordered models *à la* PS from the statistical mechanics point of view. As we started by saying, our conclusive purpose is however that of generally infer "up to which" point the disorder is relevant, hence to possibly make predictions on the expected qualitative behaviours of the main observables in real sequences of a given order of length, in the case in which one chooses the theoretically correct value of $c_p = 2.15$ for the exponent.

Let us introduce [36] a disordered SAWDNA (DSAWDNA) in which the sequence (*i.e.*, the two complementary sequences of the two strands) is chosen by assuming a binomial probability distribution for the two different kinds of possible base pairs. In this DSAWDNA model, two monomers can still occupy the same point on a $3D$ cubic lattice *if and only if* they occupy the same position i ($i \in [0, 1, 2, \ldots, \mathcal{N}]$) in the two chains (which also implies that base pair mismatches are not allowed). On the other hand, the energy gain is different between the AT and the GC base pairs, and this is taken into account by introducing chain position dependent variables $\{\epsilon_i\}$. For the sake of simplicity, these variables are assumed to be independently identically distributed random variables, that, as advanced, follow the binomial law:

$$P(\epsilon_i) = \frac{1}{2}\left[\delta(\epsilon_i - \epsilon_{AT}) + \delta(\epsilon_i - \epsilon_{GC})\right]. \tag{2.4}$$

Here one can take, in particular, $\epsilon_{GC} = 2\epsilon_{AT} = 2\epsilon$. It is to be noted that the system thermodynamic behaviour depends in any event only on the scaled variable $\epsilon\mathcal{T}$. Moreover, the chosen ratio between the linking energies reproduces its correct order of magnitude that is known from the experiments [25]. Indeed, just this last choice turned out to allow to relate the observed behaviour to the experimental one, as will be clear in the following Section 2.7.6.

Importantly, the $\{\epsilon_i\}$ are *quenched* random variable, *i.e.*, their values are fixed and they do not change during the thermal process of denaturation that one is studying. To discuss in some depth the subject of systems with quenched disorder is definitely outside the aims of the present work. In fact, it is an extremely extended subject on which there is a quite huge literature. We simply recall that the presence of quenched disorder itself is equivalent, from the theoretical point of view, to the need of calculating the average over disorder after the average over the thermal variables. More exactly, by assuming to be in the canonical ensemble for simplicity, one is interested in averaging over disorder the free energy, hence the logarithm of the partition function. This is usually a hard task that motivated the use of *ad hoc* methods, such as the replica theory, originally introduced in the spin glass research field [48].

In the present model, the single sequence Hamiltonian is given by $\mathcal{H} = -\sum_{i=1}^{\mathcal{N}} \epsilon_i \sigma_i$. The $\sigma_i \in \{0, 1\}$ are the *thermal* variables, with $\sigma_i = 0$ if the base pair in position i is open and $\sigma_i = 1$ if it is closed. Thus, differently first of all from spin glasses, in the present model there is no frustration, from the point of view that there are no competing interactions [49]. Moreover, there is clearly a single configuration of the thermal variables that corresponds to the lowest minimimum of the energy, both globally and locally, namely the one in which the considered base pairs are coupled. To make another example, that is relevant from the biomathematical point of view, the behaviour of the models for the RNA folding appears more subtle to be theoretically understood (see for instance [50, 51, 52], and references therein). In fact, in the present model, at *finite* temperature, the number of different configurations that share similar energy and entropy cannot be as large as in the RNA case. Furthermore, these configurations are more similar the ones to the others then in the RNA case, hence here one does not expect to have as much large free energy barriers to be overtaken in order to pass from a configuration to another one, or, in other words, no frustration from the topological

/ dynamical point of view (though some models for DNA denaturation can display aging behaviour [53]).

In any event, to correctly calculate the free energy is a hard task in the case of the DNA, too, and in particular the application of the replica method to disordered PS models with $c_p > 2$, in the regime of *weak* disorder, would predict a transition of the first order as in the homogeneous case [22]. On the other hand, a study in the regime of *strong* disorder, predicts the presence of two transition, with the low temperature one corresponding to a Griffiths singularity (*i.e.*, a singularity of infinite order in the free energy) [23]. In detail, we refer the reader to [24] for a comprehensive recent theoretical work on this subject, that points towards the presence of an infinite order singularity in this class of models, too. Qualitatively, we find that it could be important to better distinguish, from the theoretical point of view, between the case in which one is allowing also for repulsive link energies and the case in which all the link energies are supposed to be positive, though of different strengths, as in the present work and in the previous numerical results [36, 37, 38, 39].

On general bases (see for instance [54] and references therein), one expects that the thermodynamic limit behaviour near the critical point of a disordered model is still characterizable by critical exponents, that can or cannot be different from the ones of the same model without disorder. The disorder is irrelevant if this random critical point is characterized by the same critical exponent as the critical point of the same model without disorder, whereas the disorder is relevant in the opposite case.

In the present case, by hypothesizing the simplest possible behaviour, one can expect to find a random fixed point characterized by a single independent critical exponent as in the case of the pure model, the SAWDNA that we discussed in the previous Section 2.1. Therefore, one is led to the question if the exponent related to the (average) loop-length probability distribution, c_r, is or is not different from $c_p \simeq 2.15$ (once again, here, r refers to the random critical point, whereas p refers to the pure critical point).

Though the introduced DSAWDNA model is still a very simple model, the theoretical answer to this questions is far from being obvious. Is disorder relevant? And, if disorder is relevant, is it relevant in the usual sense, *i.e.*, does it have the effect to round-off the transition? In other words, is $c_r \neq c_p$ with, in particular, $c_r < 2$? Does the disordered model undergo a smooth transition?

2.3 Theoretical predictions

A basic Renormalization Group theoretical prediction on the cases in which one should expect the disorder relevance dates back to the work of Harris of 1974 [55]. Though the result is obtained in the context of Ising models, it turned out to be much more generally applicable. Moreover, it has the advantage that it can be very simply stated. The disorder is relevant whenever the specific heat of the pure model is diverging, *i.e.*, $\alpha_p = 2 - D\nu_p > 0$.

In detail, as partially advanced, the disorder relevance implies that the thermodynamic limit behaviour of the model is described by a *random* fixed point, that is characterized by different exponents with respect to the ones of the *pure* fixed point, that describes the thermodynamic limit behaviour of the pure system. In such cases, on the basis of the same *Harris criterion*, one can furthermore predict a smoother transition in the disordered system, with $\alpha_r = 2 - D\nu_r < 0$.

Clearly, when applying this statement to models *à la* PS with $c_p = 2.15$, that undergo a first order transition with $\alpha_p = 1$, one predicts that the disordered model has a random fixed point characterized in particular by $\alpha_r < 0$. This implies a transition smoother than the first order pure model one, with $c_r < 2$. In fact, the prediction of disorder relevance in the here considered case is suggested by other, both relevant and different, theoretical results [18, 56, 57], that studied in detail the case of a first order phase transition in the pure model, too.

Nonetheless, it was not clear, when the DSAWDNA was introduced and numerically studied in [36], if the Renormalization Group predictions [55, 18] were to be considered correctly applicable

to models *à la* PS with $c_p > 2$, since these are models of interacting polymers, that, despite the presence of a diverging correlation length, undergo a first order transition because of self-avoidance effects, hence of infinite range effects in the thermodynamic limit. On the other hand, it is also to be noted that the "almost" uni-dimensionality of the models makes the transition a peculiar first order transition, with practically zero interface contribution. Correspondingly, also the general predictions for the smoothing of first order transition in the presence of disorder [56, 57] appeared to be considered with some caution.

As we are going to discuss, apart from the subsequent mathematical results [20, 21], actually there are further important peculiarities of the models *à la* PS with $c_p > 2$, that could theoretically suggest a more subtle behaviour in the thermodynamic limit. In fact, one could expect relevance of disorder in an unusual way [38, 39]. This is the hypothesis, supported in the same works by numerical simulation findings, that motivated the present extensive analyses. Nevertheless, for the sake of clarity, let us first of all complete the sketching of the background.

2.4 Demanding numerical simulations

When numerically studying systems with quenched disorder, in order to attempt to understand the thermodynamic limit behaviour, one applies the same kind of finite size scaling techniques, that are based on the Renormalization Group [8], as in the pure case. From this point of view, we refer the reader in particular to [54], where it is exhaustively discussed the case of the diluted Ising model. Indeed, this is a quenched disordered system that is similar to the present one, notably as regards the fact that it does not display competing interactions, and accordingly no frustration.

With the aim of recalling the key ingredients of the finite size scaling techniques, let us remind that, following the Renormalization Group [8], the only relevant dimensionless ratio near the critical point is the one between the thermodynamic limit correlation length, $\xi(\mathcal{T}) \sim |\mathcal{T}_c - \mathcal{T}|^{-\nu}$, and the linear scale L of the system under consideration (namely, in the present case, $L = \mathcal{N}$). In a system without quenched disorder ($\nu = \nu_p$) this implies that an observable $\mathcal{O}$, with thermodynamic limit behaviour of its singular part described by the critical exponent e_p ($\lim_{\mathcal{T}\to\mathcal{T}_c} \mathcal{O}(\mathcal{T}) \propto |\mathcal{T}_c - \mathcal{T}|^{e_p}$), is expected to behave according to the law:

$$\mathcal{O}(L, \mathcal{T}) = L^{-e_p/\nu_p} \tilde{\mathcal{O}}[(\mathcal{T}_c - \mathcal{T})L^{1/\nu_p}], \tag{2.5}$$

with $\tilde{\mathcal{O}}$ a scaling function.

As advanced, an analogous scaling picture is expected to describe random critical points [54]. In fact, this is the theoretical basis that allows to evaluate the correlation length critical exponent ν_r, as well as the critical exponent e_r related to the considered observable. Such an evaluation is performed by studying the scaling behaviour of $\overline{\mathcal{O}(\{\epsilon_i\}, \mathcal{N}, \mathcal{T})}$. Here, $\overline{(\cdot)}$ means the *standard* average over the quenched variables $\{\epsilon_i\}$, *i.e.*:

$$\overline{\mathcal{O}(\{\epsilon_i\}, \mathcal{N}, \mathcal{T})} = \frac{1}{N_s} \sum_{\{\epsilon_i\}_h, h=1}^{N_s} \mathcal{O}(\{\epsilon_i\}_h, \mathcal{N}, \mathcal{T}), \tag{2.6}$$

where N_s is the total number of considered disorder configurations (random sequences) for the given chain length $\mathcal{N}$ and $\mathcal{O}(\{\epsilon_i\}_h, \mathcal{N}, \mathcal{T})$ is the measured value of the observable for h-th disorder configuration, $\{\epsilon_i\}_h$, at the given temperature $\mathcal{T}$. Importantly, this average is not to be confused with the *non standard* one that we will introduce subsequently. The difference is that, usually, as evident from (2.6), the average at a given temperature is taken among the values of the observable at that temperature, independently of the disorder configuration.

In any event, already in the case of the standard way of averaging, for systems with quenched disorder, the finite size scaling techniques require to collect data on each studied observable, at each

temperature of the chosen set, for a sufficiently large number N_s of different disorder configurations. Moreover, the whole procedure needs to be repeated up to sufficiently large sizes (chain lengths).

We do not enter, here, into the details of the numerical simulations on the DSAWDNA in [36], that, as in the ordered SAWDNA case, could be performed thanks to an appropriate algorithm, *i.e.*, the Pruned-Enriched Rosenbluth Method (PERM) [58], and that moreover required a quite massive use of a powerful parallel computer. We limit ourselves to note that the sequences of length $\mathcal{N}_b > \mathcal{N}_a$ were obtained by adding $\mathcal{N}_b - \mathcal{N}_a$ randomly generated variables to the sequence of length $\mathcal{N}_a$. As discussed in [32], this introduces some correlations, though it is not expected to influence the evaluation of the critical exponent. Despite this fact, and despite that the simulations were performed in parallel, it was possible to collect a reasonably large statistics only up to chains of length $\mathcal{N} = 800$. Indeed, as we will better clarify in the following Section 2.7.6, for the considered model, such an order of magnitude of $\mathcal{N}$ is too small for correctly extrapolating the thermodynamic limit behaviour. In fact, the best possible evaluations agreed with values of c_p smaller than 2, but anyway corresponding to a divergent specific heat. Furthermore, on the basis of the collected data, one could not rule out the possibility of a first order transition, also in presence of disorder.

2.5 On the possibility of a peculiar behaviour

Briefly, not only the numerical results on the DSAWDNA did not answer to the question if the exponent c_r were different from c_p. Because of the previously recalled general expectations on the effect of disorder, they especially suggested that this model could display a peculiar behaviour, *i.e.*, a first order transition as in the pure case despite $\alpha_p > 0$, or, in any event, a smoother transition than in the pure case but characterized by $\alpha_r > 0$.

Then, on general bases, one would expect that such a possibly peculiar behaviour could concern the whole universality class of the models *à la* PS with $c_p > 2$, that undergo a first order transition in the thermodynamic limit in the pure case, and in which, therefore, the self-avoidance is more or less *globally* taken into account. Moreover, as already outlined, that a correct model for DNA denaturation could undergo a first order transition in the presence of disorder, too, can appear reasonable from the point of view of the experimental results. In fact, the hypothesis of a possible peculiar behaviour of these models can be proposed on grounded theoretical bases [38, 39], and there is evidence for a similar peculiar behaviour in different cases (see, for instance, [59] and [60]).

In order to sketch the theoretical background of such a hypothesis, let us underline, first of all, that the models *à la* PS with $c_p > 2$ share the presence of an infinite range interaction, the self-avoidance itself. This makes not evident *a priori*, in particular in the disordered case, that one could apply the standard argument used to demonstrate self-averageness of densities of extensive quantities [61], that is built on the description of the system as consisting of weakly interacting sub-systems. Especially, this argument breaks out, in any event, at the critical point, because of the presence of a diverging correlation length. It is worth noticing that the first order transition in the considered class of pure models is, just for this reason, a peculiar first order transition, *i.e.*, in the usual first order transitions there is not an "almost" dimensionless interface and, correspondingly, the correlation length does not diverge.

In fact, it was generally put forward [62, 63] that, when considering random critical points, results of standard finite size scaling analyses should be considered with some care. Within this framework, following a general Renormalization Group result for random ferromagnets [19], we are led to introduce an additional observable, the effective critical temperature (or pseudo-critical temperature), $\mathcal{T}_c(\{\epsilon_i\}_h, \mathcal{N})$, and to consider its dependence on the disorder configurations. In the present "almost" uni-dimensional case, with $L = \mathcal{N}$, the mean value of this quantity:

$$\mathcal{T}_c(\mathcal{N}) \equiv \overline{\mathcal{T}_c(\{\epsilon_i\}, \mathcal{N})}, \tag{2.7}$$

and its fluctuations:

$$\delta\mathcal{T}_c(\mathcal{N}) \equiv \left\{ \overline{\left[\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})\right]^2} - \left[\overline{\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})}\right]^2 \right\}^{1/2}, \tag{2.8}$$

are expected to behave, as functions of the system size [19, 62, 63], as:

$$\left.\begin{array}{lcl} \mathcal{T}_c(\mathcal{N}) & \simeq & \mathcal{T}_c + C\mathcal{N}^{-1/\nu_p} \\ \delta\mathcal{T}_c(\mathcal{N}) & \propto & \mathcal{N}^{-1/2} \end{array}\right\} \text{ for irrelevant disorder} \tag{2.9}$$

$$\left.\begin{array}{lcl} \mathcal{T}_c(\mathcal{N}) & \simeq & \mathcal{T}_c + C\mathcal{N}^{-1/\nu_r} \\ \delta\mathcal{T}_c(\mathcal{N}) & \propto & \mathcal{N}^{-1/\nu_r} \end{array}\right\} \text{ for relevant disorder} \tag{2.10}$$

Noticeably, the theoretical framework [19] predicts that, for relevant disorder, the mean value and the fluctuations of the pseudo-$\mathcal{T}_c$ should scale with the same exponent. In fact, this is the same framework [55] that allows to infer disorder relevance as soon as the specific heat critical exponent for the pure system fulfills the condition $\alpha_p > 0$. The case in which the mean value and the fluctuations of the pseudo-$\mathcal{T}_c$ are described by two different exponents (see (2.9)), corresponds instead to the situation that is usually encountered when disorder is irrelevant (as the specific heat of the pure system displays no divergence, *i.e.*, $\alpha_p < 0$).

In fact, it is this last case that can be also attributed to the presence of two correlation lengths. Importantly, the theoretical basis for such a possibility was laid once again within the Renormalization Group framework, in particular within the context of the random transverse field Ising chains [64, 65]. This is the scenario that was proposed to apply to the class of models *à la* PS with $c_p > 2$, too, in [38, 39]. More in detail, here, an independently diverging correlation length can be pictured as related to the free-end distance, or, possibly on more grounded bases, it can be hypothesized that the divergence of the *typical* loop is different from that of the *average* one. Accordingly, the expected transition in the disordered models is of first order, as in the pure case, from the point of view of the behaviour of the *typical* observables, *i.e.*, the given sequence undergoes a first order transition, with $\nu_{r,1} = 1$. Nonetheless, one expects a second order transition from the point of view of the *average* observables, whose behaviour is ruled by $\nu_{r,2} = 2$. This scenario appears furthermore supported, since numerical evidence for such a peculiar behaviour was found in the case of a disordered PS model, with $c_p = 2.15$, in the same works [38, 39], in which the model is studied up to chains of length as large as $\mathcal{N} = 2 \cdot 10^6$.

The proposed scenario also implies [62, 63] that the standard scaling law, that describes the finite size behaviour of a thermodynamic observable $\mathcal{O}$ with critical exponent e_r (see (2.5)), is expected to be better obeyed by the quantity $\overline{\overline{\mathcal{O}}}$:

$$\overline{\overline{\mathcal{O}(\{\epsilon_i\},\mathcal{N},\mathcal{T})}} = \mathcal{N}^{-e_r/\nu_r}\tilde{\mathcal{O}}\left[(\mathcal{T}_c(\mathcal{N}) - \mathcal{T})\mathcal{N}^{1/\nu_r}\right], \tag{2.11}$$

in which we label by $\overline{\overline{(\cdot)}}$ the average over disorder performed by taking into account the sequence-dependent $\mathcal{T}_c(\{\epsilon_i\}_h,\mathcal{N})$, according to:

$$\overline{\overline{\mathcal{O}(\{\epsilon_i\},\mathcal{N},\mathcal{T})}} = \frac{1}{N_s}\sum_{\{\epsilon_i\}_h h=1}^{N_s} \mathcal{O}\left\{\{\epsilon_i\}_h,\mathcal{N},[\mathcal{T} - \mathcal{T}_c(\{\epsilon_i\}_h,\mathcal{N}) + \mathcal{T}_c(\mathcal{N})]\right\} \tag{2.12}$$

In such a way, one avoids the results being ruled by the fluctuations of the pseudo-$\mathcal{T}_c$.

Finally, in the case of disorder relevance, theoretical results [19] imply *strong* non self-averageness in the thermodynamic observables that are *singular* at the critical point, despite that these are densities of extensive quantities. By definition, self-averageness is measured from the ratio:

$$\mathcal{R}_{\mathcal{O}} = \frac{\overline{\left[\mathcal{O}(\{\epsilon_i\},\mathcal{N},\mathcal{T})\right]^2} - \left[\overline{\mathcal{O}(\{\epsilon_i\},\mathcal{N},\mathcal{T})}\right]^2}{\left[\overline{\mathcal{O}(\{\epsilon_i\},\mathcal{N},\mathcal{T})}\right]^2}, \tag{2.13}$$

with the strong non self-averaging behaviour corresponding to $\mathcal{R}_{\mathcal{O}} \sim 1$ (whereas $\mathcal{R}_{\mathcal{O}} \sim 1/\mathcal{N}$ in the usual self-averaging behaviour). Importantly, in the present case, these results for observables such as the order parameter and the susceptibility (that here behaves as the specific heat) are expected both in the case in which the mean value and the fluctuations of the pseudo-critical temperature scale with the same exponent and in the case of the peculiar behaviour that corresponds to a pseudo first order transition.

2.6 The considered PS model with $c_p = 2.15$

The difficulty in reaching chain lengths that are larger than $\mathcal{N} = 800$, in the DSAWDNA, can be overtaken by studying the corresponding off-lattice PS model with $c_p = 2.15$. We moreover selected the other appropriate parameters, so that one exactly reproduces the observed on-lattice model behaviour, as numerically checked in [37].

Off-lattice, the interesting observables can be measured by recurrently computing the partition function. Let us recall in detail the procedure, as reported in [37, 32], for a chain of length $\mathcal{N}$ and a given sequence $\{\epsilon_i\}$. The sequence is extracted at the beginning, in agreement with the binomial distribution (2.4), with $\epsilon_{GC} = 2\epsilon_{AT}$. The only other parameter to be fixed is the value of the connectivity constant μ: in fact, $\mu = 2D$ for a D-dimensional RW on a hyper-cubic lattice and $\mu \simeq 4.7$ for a $3D$ SAW on a cubic lattice, leading to $\log \mu \simeq 1.54$ in our case [37].

For a system of length $n+1$, with both base pairs $i = 1$ and $i = n+1$ in the closed state ($\sigma_1 = \sigma_{n+1} = 1$), the *forward* partition function $Z^f(\{\epsilon_i\}, n+1, \mathcal{T})$, that sums the contributions of all the configurations, that are weighted by their Boltzmann factors, is obtained from $Z^f(\{\epsilon_i\}, n, \mathcal{T})$ as:

$$Z^f(\{\epsilon_i\}, n+1, \mathcal{T}) = e^{(\beta\epsilon_{n+1} - \log\mu)} \left\{ Z^f(\{\epsilon_i\}, n, \mathcal{T}) + \sum_{n'=1}^{n-1} \frac{Z^f(\{\epsilon_i\}, n', \mathcal{T})}{[2(n-n'+1)]^{c_p}} \right\}. \tag{2.14}$$

Similarly, the *backward* partition function $Z^b(\{\epsilon_i\}, n-1, \mathcal{T})$ is obtained as:

$$Z^b(\{\epsilon_i\}, n-1, \mathcal{T}) = e^{(\beta\epsilon_{n-1} - \log\mu)} \left\{ Z^b(\{\epsilon_i\}, n, \mathcal{T}) + \sum_{n'=n+1}^{\mathcal{N}} \frac{Z^b(\{\epsilon_i\}, n', \mathcal{T})}{[2(n'-n+1)]^{c_p}} + 1 \right\}. \tag{2.15}$$

Here, the last term in the expression between curled brackets takes into account the allowance for the free ends.

The total partition function for a sequence $\{\epsilon_i\}$ of length $\mathcal{N}$ is given by:

$$Z(\{\epsilon_i\}, \mathcal{N}, \mathcal{T}) = \sum_{n=1}^{\mathcal{N}} Z^f(\{\epsilon_i\}, n, \mathcal{T}) = Z^b(\{\epsilon_i\}, 1, \mathcal{T}). \tag{2.16}$$

Therefore, it is possible to evaluate the probability for a base pair in position n of being in the closed state (*i.e.*, the thermal average $\langle\sigma_i\rangle$):

$$\mathcal{P}(\{\epsilon_i\}, \mathcal{N}, \mathcal{T}, n) = \langle\sigma_i\rangle = \frac{Z^f(\{\epsilon_i\}, n, \mathcal{T}) Z^b(\{\epsilon_i\}, n, \mathcal{T})}{Z(\{\epsilon_i\}, \mathcal{N}, \mathcal{T}) e^{(\beta\epsilon_n - \log\mu)}}, \tag{2.17}$$

from which relevant thermodynamic observables (whose definitions are recalled in [Tab. 2.1]) are easily obtained. It is worth noticing that results in [36, 37] are in agreement with a situation in which, also in the presence of disorder, the order parameter behaves as the energy (with $\beta_r = \nu_r - 1$), and the susceptibility as the specific heat (with $\gamma_r = \alpha_r = 2 - \nu_r$). Under such conditions, it appears reasonable to expect that only one independent critical exponent needs to be evaluated.

$e(\{\epsilon_i\},\mathcal{N},\mathcal{T})$	$\equiv -$	$\frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}}\epsilon_i\langle\sigma_i\rangle$	(energy density)
$c(\{\epsilon_i\},\mathcal{N},\mathcal{T})$	$\equiv$	$\frac{1}{\mathcal{T}^2}\frac{\partial}{\partial\mathcal{T}}e(\{\epsilon_i\},\mathcal{N},\mathcal{T})$	(specific heat)
$\Theta(\{\epsilon_i\},\mathcal{N},\mathcal{T})$	$\equiv$	$\frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}}\langle\sigma_i\rangle$	(order parameter)
$\chi(\{\epsilon_i\},\mathcal{N},\mathcal{T})$	$\equiv$	$\begin{cases} \frac{1}{\mathcal{N}}\left[\langle\left(\sum_{i=1}^{\mathcal{N}}\sigma_i\right)^2\rangle-\left(\langle\sum_{i=1}^{\mathcal{N}}\sigma_i\rangle\right)^2\right] \\ \mathcal{T}\left[\frac{\partial}{\partial\epsilon_{AT}}\Theta(\{\epsilon_i\},\mathcal{N},\mathcal{T})+\frac{\partial}{\partial\epsilon_{GC}}\Theta(\{\epsilon_i\},\mathcal{N},\mathcal{T})\right] \end{cases}$	(susceptibility)

Table 2.1: Definitions of the relevant thermodynamic observables in the model (we label by $\langle\cdot\rangle$ the thermal averages).

Finally, we recall that the loop-length probability distribution, apart from the normalization constant (to be chosen by imposing $\sum_l P(\{\epsilon_i\},\mathcal{T},\mathcal{N},l)=1$), can be evaluated as:

$$P(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)\propto\frac{1}{l^{c_p}}\sum_n\frac{Z^f(\{\epsilon_i\},n,\mathcal{T})Z^b(\{\epsilon_i\},n+l+1,\mathcal{T})}{Z(\{\epsilon_i\},\mathcal{N},\mathcal{T})}. \tag{2.18}$$

This expression further highlights the fact that, in the present model, c_p corresponds to an *input*.

In this context, the simplest picture describing the random fixed point, in agreement with the numerical results in [36, 37], is the one in which in the thermodynamic limit:

$$P(\mathcal{T},l)=\lim_{\mathcal{N}\to\infty}\overline{P(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)}\propto\frac{e^{-l/\xi(\mathcal{T})}}{l^{c_r}}, \tag{2.19}$$

Accordingly, as in the pure case (see (2.18)), the average loop-length probability distribution is expected to display a purely algebraic decay at the critical point, where the average correlation length $\xi(\mathcal{T})$ diverges. This power law decay is described by the random critical exponent c_r, that, in the case of a smooth transition, is expected to be linked to the correlation length critical exponent ν_r by the relation $c_r=1+1/\nu_r$: the same kind of relation that is known to be valid for the pure system, for $c_p<2$.

In [37], in collaboration with E. Yeramian, we performed numerical simulations of such a PS model up to chain lengths $\mathcal{N}=20000$. As we will explain in detail in the following Section 2.7.1, this was made possible by the SIMulation with EXponential (SIMEX) algorithm, in which one approximates power laws (here $1/l^{c_p}$) with sums of exponentials. In order to better be able to compare the results, we applied the algorithm by choosing exactly the same values for the parameters in the approximation as in [38, 39]. On the other hand, it is in our work that the algorithm is used for obtaining data to be extensively analyzed by standard finite size scaling techniques, for the first time to our knowledge.

One could wonder if the considered numerical approximation is correct within this context, since it clearly would change the thermodynamic limit model behaviour from the analytical point of view. Indeed, the SIMEX appears a valid approximation since, when performing a finite size analysis, one is only interested in the behaviour of the observables up to the given maximum size, and the same approximation was already found to give a behaviour of $1/l^{c_p}$ that is practically indistinguishable,

within the computer numerical precision, from the analytical behaviour up to l values as large as $l = 2 \cdot 10^6$ in [38, 39].

Briefly resuming the main results in [37], in that work we performed a standard finite size scaling analysis, in which the observables were averaged according to (2.5). On the one hand, as advanced, we found numerical evidence for a model behaviour in very good agreement with the one observed in [36] up to the there considered, much smaller, maximum chain length, $\mathcal{N} = 800$. On the other hand, the reached maximum chain length, $\mathcal{N} = 20000$, turned out to be large enough for clearly extrapolating the thermodynamic limit model behaviour, despite the presence of strong finite size corrections. With the considered standard finite size scaling analysis, by studying the scaling of the order parameter and of the specific heat, we found in particular $\nu_r \sim 3$, in agreement with the best possible independent evaluation of $c_r \leq 1.5$ obtained from the data on the average probability distribution of the loop length.

These results were in striking contrast with the ones previously obtained in [38, 39], where the observed model behaviour supports the theoretical hypothesis of a pseudo first order transition, as we already recalled. We will better clarify in the following Section 2.7.6 that, in fact, the PS model that is considered in these works differs from the present one in particular for the choice of the parameters. Nevertheless, the two PS models share the same value of the pure system loop-length probability exponent $c_p = 2.15$, that is expected to be the only relevant quantity from the point of view of the behaviour in the thermodynamic limit. On the other hand, our numerical results on a smooth transition were in agreement with the (at the time very recent) mathematical ones [20, 21], that prove relevance of disorder, without divergence of the specific heat in the random case, for a class of polymer models that should encompass the PS model with $c_p > 2$.

Therefore, the contradiction between the numerical results in [37] and the ones in [38, 39], as well as the strong finite size corrections to scaling observed in [36, 37], suggested a deeper study, that is one of the original subjects of this thesis, whose findings have been partially presented in [32], a work in collaboration with E. Yeramian. Clearly, one is first of all interested in testing the alternative hypothesis that the model could undergo a pseudo first order transition.

2.7 Conclusive numerical results

2.7.1 The SIMEX approximation and the other numerical details

As anticipated, for efficient numerical implementation, the recursive equations for the forward and backward partition functions (see (2.14) and (2.15)) in the PS model are solved with the SIMEX algorithm, relying upon the approximation of the power law $1/l^{c_p}$ with a sum of exponentials.

Let us furthermore recall that the basic idea in the SIMEX algorithm, as used in the context here [32, 37, 38, 39], was originally expressed, in [35], specifically for the numerical study of PS models of denaturation transitions in linear DNA molecules. However, the generality of the powerful idea at the basis of this representation of the long-range effect as a sum of exponentials was not appreciated to its fair value, since in the original work it was implemented in the context of conditional probabilities specific to the considered model.

Then, the formulation of the SIMEX scheme proceeded in two steps. First the original idea was reformulated in more general terms for the linear PS model, in the context of recursions written directly in terms of partition functions [67]. On such a basis, it was later possible to propose generalizations of the idea to higher order models involving several mutually coupled long-range effects [68], with the corresponding algorithmic complexities reduced by several orders of magnitude. In order to be efficient, the method relies on the representations of long-range effects as sums of exponentials, that need to be as much accurate as possible. Within this context, the Padé-Laplace method [33, 34] provides an elegant analytical solution to the problem, not only in the case of purely decaying functions, such as the power-laws, that are represented as sums of real exponentials, but more generally

allowing to represent functions with complex exponentials. As an illustration for potential applications, the SIMEX scheme was used for example to implement sequence alignments with realistic gap models, in Bioinformatics [69].

In this background, we use here the approximation:

$$\frac{1}{(2l)^{2.15}} \simeq \sum_{k=1}^{N_E} a_k \, e^{-2lb_k}, \tag{2.20}$$

with $N_E = 15$ exponential terms and the values of the coefficients $\{a_k, b_k\}$ (the same as in [38, 39]) reported in Appendix A.1, together with the implementation details, that rely upon an *ad hoc* introduction of *free-energy*-like quantities, originally proposed in [66].

In this direction, because of the importance of the underlying numerical problem, it is relevant to further recast the obtention of multi-exponential representations in the general context of approximation problems. Since Prony (1795), the problem of obtaining numerical representations of given functions as sums of exponentials has been addressed in many fields and contexts. As a matter of fact, this problem can be posed from two very different -in principle- point of views: identification or approximation. In the identification case the given function is supposed, by essence, to be a sum of exponentials and the problem consists in retrieving precisely the *genuine* number of exponential components, with the evaluations of the associated parameters. This problem is considered difficult, notably in the case of real exponentials, because of ill-conditioning. It is then easy to *over-fit* the data, with methods such as least-squares, with increasing number of components, thus missing the aim of proper identification of the model in terms of its original components. In the alternative case, related to approximation, the problem is of course completely different, as the given function, known analytically, is not a sum of exponentials and the aim is to obtain indeed the best possible approximation in such a form. The present situation, concerning the power-law for the long-range effect, is of course relevant to the approximation case, and we need to ensure an accurate representation of the power-law up to the largest considered sizes for the system with the sum of exponentials. The number of components in the multi-exponential representation will then of course depend on the maximal size of the system considered, with the need to introduce, according to this size, additional components with progressively smaller b_k parameters (see (2.20)): in the numerical fit, for increasingly larger values of the variable l, close to the maximal one in the model, the exponential components with the smallest b_k parameters are required to decay according to the corresponding values of the power-law.

In such context, the Padé-Laplace method (that encompasses various other formulations such as the Prony method or the method of moments as particular cases) was originally formulated for the identification problem. It was however also used in the approximation context, for the obtention of numerical approximations of the power-law with sums of exponentials. Interestingly, in such a case, an *identification-like* behaviour was observed, concerning the number of exponential components in the representation: more precisely, it was observed that the appropriate number of components for systems of size $\mathcal{N}_{max}$ followed essentially a law in $\ln(\mathcal{N}_{max})$ (10 components for $\mathcal{N}_{max}$ up to $2 \cdot 10^4$, 14-15 components for $\mathcal{N}_{max}$ up to 10^6).

In various studies, including in the original Fixman-Freire paper [35], multi-exponential representations for the power-law were obtained resorting to different methods. It is however interesting to underline that increasingly more complex models implemented in the context of the generalizations of the SIMEX could involve non purely decaying general long-range effects. In such case it would then be necessary to resort to approximations with sums of general complex exponentials, as allowed by the Padé-Laplace method [33, 34].

In the present work, we are referring to the SIMEX algorithm as consisting of both the approximation (2.20) and the recursive writing of the partition function. It is worth noticing that such an algorithm allows to gain a factor larger than $O(\mathcal{N}^2)$ in the numerical computation time, that

$\mathcal{N}$	N_S	$\mathcal{T}_{min}$	$\mathcal{T}_{max}$
100	2000	0.95	1.2
200	2000	0.95	1.2
500	2000	1.0	1.16
750	1000	1.0	1.16
1000	1000	1.0	1.15
2500	1000	1.02	1.14
5000	1000	1.02	1.14
7500	1000	1.04	1.12
10000	600	1.04	1.14
15000	500	1.04	1.12
20000	500	1.04	1.12

Table 2.2: Number of sequences (N_S) and temperature intervals ($[\mathcal{T}_{min}, \mathcal{T}_{max}]$) adopted in the computations, following the chain lengths ($\mathcal{N}$).

decreases from $O(\mathcal{N}^3)$ to $O(\mathcal{N})$. The order of magnitude of the reduction makes in principle possible to deal with the numerical study of more complex cases, such as the one of RNA folding, where once again one has to face power laws. Nevertheless, despite the interest of the subject, the difficulty is that the exponent is not expected to be loop-independent, for a correct modelling in which the self-avoidance is completely taken into account [51]. In any event, it is this reduction that allowed us to study in particular the detailed temperature dependence of the various observables, hence to perform the here discussed analysis.

Summarizing, in the present work (as in [37]), we study the disordered PS model with $c_p = 2.15$ by adopting the values $\epsilon_{GC}/\epsilon_{AT} = 2$ and $\log \mu = 1.54$. Without any loss of generality, we set in all computations $\epsilon_{AT} = 1$, *i.e.*, the temperature is in ϵ_{AT} unities.

The conditions chosen to collect statistics for each chain length $\mathcal{N}$ are recalled in [Tab. 2.2] ([37]), in terms of number N_S of studied different sequences $\{\epsilon_i\}$ and considered range of temperature intervals $[\mathcal{T}_{min}, \mathcal{T}_{max}]$. We select these intervals in such a way that they are roughly centered on the temperature at which the ($\mathcal{N}$-dependent) average specific heat reaches its maximum. We outline that, moreover, the temperature intervals are always divided into the (quite large) number of $N_T = 250$ equally spaced sub-intervals, which is necessary for the kind of analysis we are interested in.

In fact, for each sequence, the specific heat is evaluated by computing numerically the derivative of the energy density with $\delta\mathcal{T} = (\mathcal{T}_{max} - \mathcal{T}_{min})/\mathcal{N}_T$. On the other hand, the evaluation of the susceptibility is obtained by numerical partial derivations with respect to ϵ_{AT} and ϵ_{GC} (see [Tab. 2.1]), and the values $\delta\epsilon_{AT} = \beta\ 10^{-4}$, $\delta\epsilon_{GC} = \beta\ 10^{-4}\epsilon_{GC}$ are adopted, respectively. We checked that such choices ensure appropriate accuracy for the computations (with the errors on the positions of maxima consistently smaller than the sample-to-sample fluctuations) [37]. Let us notice, too, that the present evaluation of this last observable implies that, for each chain length, sequence and temperature, the whole procedure has to be repeated three times.

Finally, the errors on average quantities are evaluated from sample-to-sample fluctuations (*i.e.*, in the present case, the fluctuations between different sequences).

2.7.2 The definition of the pseudo-$\mathcal{T}_c$

As stated in the introductive Sections, the alternative scenario that we are interested in testing is the one in which the considered PS model undergoes a pseudo first order phase transition in the thermodynamic limit. In this case, the behaviour of the typical observables (*i.e.*, the ones measured in the study of a single sequence) is expected to be different from the behaviour of the observables

averaged in the standard way. In detail, the two behaviours are expected to be ruled by different correlation length critical exponents, $\nu_{r,1} = \nu_p = 1$ and $\nu_{r,2} = 2$. Therefore, the typical observables behave as in the pure model, that undergoes a first order phase transition, whereas the behaviour of the average observables is ruled by the fluctuations of the pseudo-critical temperature. This is the picture that has been proposed in [38, 39], and that is supported by the numerical data on the PS model with $c_p = 2.15$ that is studied in these works.

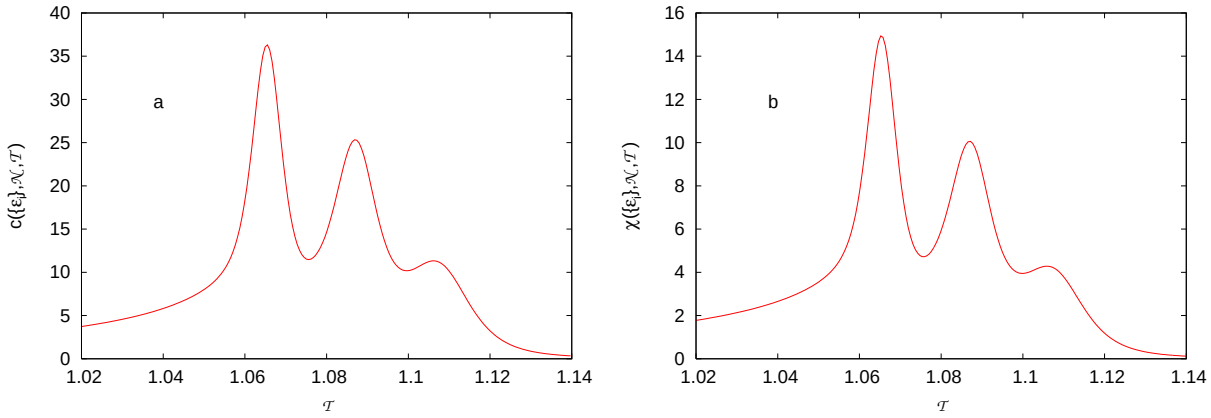

Figure 2.6: **a) Specific heat c; b) susceptibility χ.** Data are for a given sequence of length $\mathcal{N} = 2500$, plotted as function of the temperature, $\mathcal{T}$.

It is worth noticing from the beginning that there is an evident difference between the numerical results in [38, 39] and the ones both on the DSAWDNA numerically studied in [36] and on the present PS model, that was proposed and initially numerically studied in [37]. This difference regards in particular the behaviour of the order parameter, the density of closed base pairs $\Theta(\{\epsilon_i\},\mathcal{N},\mathcal{T})$, whose definition is given in [Tab. 2.1]. Indeed, in [38, 39], the single sequence order parameter behaves much as in the pure model, by varying quite abruptly from zero to nearly one at a given temperature, that appears strongly sequence dependent. On the other hand, both in [36] and in [37, 32] it is observed, for a fraction of the sequences that appears to depend on the considered chain length, a multi-step behaviour. Correspondingly, it is observed a multi-peak behaviour both in the given sequence specific heat $c(\{\epsilon_i\},\mathcal{N},\mathcal{T})$ and in the given sequence susceptibility $\chi(\{\epsilon_i\},\mathcal{N},\mathcal{T})$, whose definitions are given in [Tab. 2.1], too.

For the sake of clarity, we report in [Fig. 2.6] the data on these last two quantities, as functions of the temperature $\mathcal{T}$, for a given sequence of length $\mathcal{N} = 2500$, whose order parameter displays three well distinguishable steps: both of them display three distinct peaks. It is first of all to be outlined that the behaviours observed in the present case appear to support the hypothesis that the considered observables are non self-averaging (see (2.13)), *i.e.*, $\mathcal{R}_\Theta \sim \mathcal{R}_c \sim \mathcal{R}_\chi \sim 1$, without the need of further analysis. Nonetheless, the non self-averageness of the order parameter is a feature shared with the slightly different PS model studied in [38, 39], just because of the strong sequence dependence of the temperature at which one observes its abrupt variation.

Most importantly, the absence / presence of multi-step behaviour in the order parameter already suggests that, despite the same $c_p = 2.15$ value, the present PS model and the one considered in [38, 39] substantially do not display the same qualitative behaviour. In fact, as we will explain, this could depend on a huge difference in the order of magnitude of the chain length "at which" the disorder relevance becomes evident.

In any event, the different observed behaviours makes questionable, in our case, the use of the same definition for the pseudo-critical temperature itself. In fact, in the present work, the pseudo-$\mathcal{T}_c$ is introduced and studied in the case of a system in which the single sample order parameter displays multi-steps (while the single sample susceptibility displays multi-peaks) as function of the

temperature for the first time to our knowledge. In the works in [62, 63], the pseudo-$\mathcal{T}_c$ was defined as the temperature that corresponds to the unique maximum of the single sample susceptibility. In particular on this background, we found appropriate to define the pseudo-critical temperature, $\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})$, for a given sequence $\{\epsilon_i\}$ of length $\mathcal{N}$, *as the temperature at which the single sequence susceptibility reaches its global maximum*. We underline that this temperature is very close to the temperature at which the specific heat reaches its global maximum, too, since these observables always display a similar behaviour, as in the data shown in [Fig. 2.6].

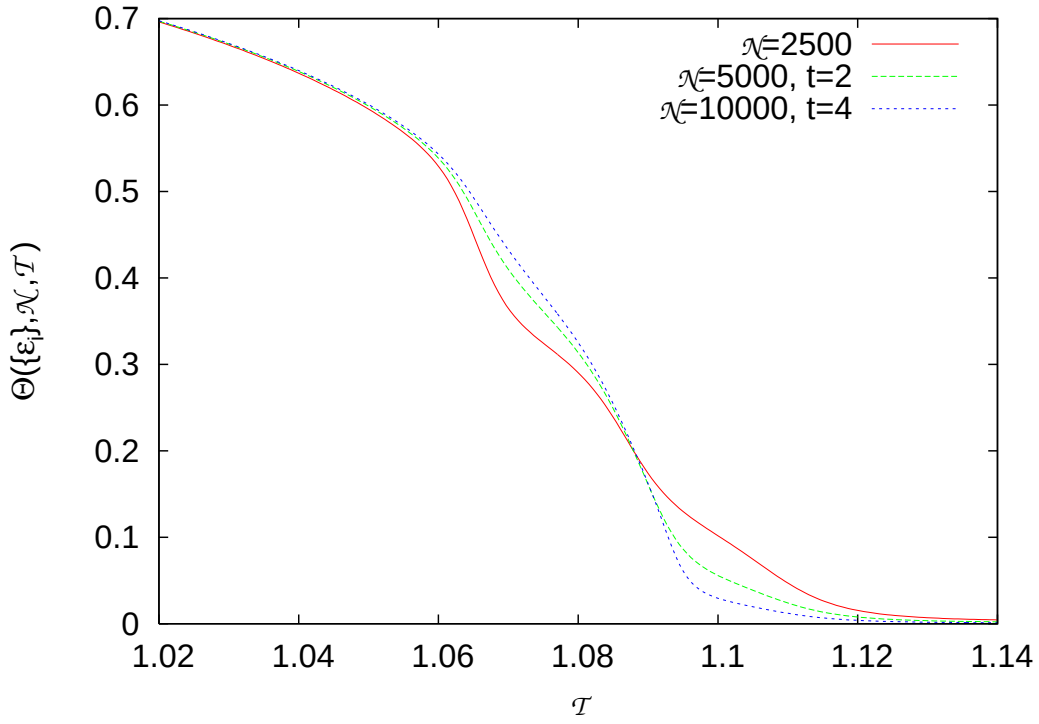

Figure 2.7: **Order parameter.** Here, Θ is plotted as function of the temperature for a given sequence of length $\mathcal{N} = 2500$ (the same sequence as in [Fig. 2.6] and for the sequences of lengths $\mathcal{N} = 5000$ and $\mathcal{N} = 10000$ obtained by concatenating $t = 2$ and $t = 4$, respectively, copies of the original sequence.

On the other hand, in [38, 39], two different definitions were considered: one of these was obtained on the basis of the free-energy behaviour, whereas in the other the pseudo-critical temperature $\mathcal{T}_c'(\{\epsilon_i\},\mathcal{N})$, for a given sequence $\{\epsilon_i\}$ of length $\mathcal{N}$, was simply related to the crossing point among the order parameters $\Theta(\{\epsilon_i\},\mathcal{N} = \mathcal{N}_0,\mathcal{T})$, $\Theta(\{\epsilon_i\},\mathcal{N} = 2\mathcal{N}_0,\mathcal{T})$ and $\Theta(\{\epsilon_i\},\mathcal{N} = 4\mathcal{N}_0,\mathcal{T})$. Here, the sequences of length $2\mathcal{N}_0$ and $4\mathcal{N}_0$ corresponded to the concatenation of $t = 2$ and $t = 4$, copies of the original sequence of length $\mathcal{N}$, respectively.

In [Fig. 2.7] we report the data on the order parameter as function of the temperature for $\mathcal{N} = 2500$, in the case of the same sequence as for the specific heat and the susceptibility reported in [Fig. 2.6], that, as advanced, displays three well distinguishable steps. We also plot the behaviour of this quantity for the sequences of length $2\mathcal{N}_0 = 5000$ and $4\mathcal{N}_0 = 10000$, respectively, that are obtained by the concatenation procedure.

Despite the multi-step behaviour, the data for the original sequence and the ones obtained in such a way display the same intersection point. Therefore, it is in principle possible to adopt the same simple definition of the pseudo-$\mathcal{T}_c$ as in [38, 39] in our case, too. Indeed, we checked, up to the relatively large size of $\mathcal{N} = 2500$, that this different definition gives compatible results with the one that we propose.

It is worth underlining that our definition of the pseudo-$\mathcal{T}_c$ as the abscissa of the global maximum

of the susceptibility is made possible by the accurately measured temperature dependence of the observables in the present study. It is moreover interesting to notice that this definition appears the most reasonable one in particular since, with this choice, by applying (2.11), one has:

$$\begin{aligned}\max_{\mathcal{T}} \overline{\overline{\chi(\{\epsilon_i\},\mathcal{N},\mathcal{T})}} &= \max_{\mathcal{T}} \frac{1}{N_s} \sum_{\{\epsilon_i\}_h,h=1}^{N_s} \chi\left\{\{\epsilon_i\},\mathcal{N},[\mathcal{T}_c(\mathcal{N})-\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})+\mathcal{T}]\right\} = \\ &= \frac{1}{N_s} \sum_{\{\epsilon_i\}_h,h=1}^{N_s} \chi\left[\{\epsilon_i\},\mathcal{N},\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})\right] = \overline{\max_{\mathcal{T}} \chi(\{\epsilon_i\},\mathcal{N},\mathcal{T})}. \qquad (2.21)\end{aligned}$$

Accordingly, such a definition should be especially effective for providing evidence for a possible linearly diverging behaviour of the typical susceptibility (*i.e.*, in the present case in which the susceptibility behaves as the specific heat, for a possible specific heat exponent $\alpha_{r,1} = 1 = \alpha_p$, and also, by hyper-scaling in $D = 1$, for $\nu_{r,1} = 1 = \nu_p$).

2.7.3 The scaling behaviour of the pseudo-$\mathcal{T}_c$ and of its fluctuations

The values of $\mathcal{T}_c(\mathcal{N}) = \overline{\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})}$ and $\delta\mathcal{T}_c(\mathcal{N}) = \{\overline{[\mathcal{T}_c(\mathcal{N})-\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})]^2}\}^{1/2}$, as functions of $1/\mathcal{N}$, are plotted in [Fig. 2.8], along with the best fits obtained following $\mathcal{T}_c(\mathcal{N}) \propto \mathcal{T}_c + C\,\mathcal{N}^{-1/\nu_{r,1}}$ and $\delta\mathcal{T}_c(\mathcal{N}) \propto \mathcal{N}^{-1/\nu_{r,2}}$, hence allowing the possibility, in principle, of two different exponents (see (2.9)-(2.10) and the associated discussions). In addition, data for the mean value and for the fluctuations of $\mathcal{T}_c'(\{\epsilon_i\},\mathcal{N})$, that is defined according to [38, 39], are also plotted in [Fig. 2.8] for $\mathcal{N} \leq 2500$. For each $\mathcal{N}$-value, we moreover checked that the behaviour of $\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})$ agrees with a Gaussian distribution. We notice first of all that, interestingly, whereas strong finite size corrections are observed in the behaviour of average quantities when applying standard scaling laws [37], it appears that the data on the mean value and on the fluctuations of the pseudo-critical temperature agree well with the corresponding expected laws on the whole $\mathcal{N}$-range considered.

On such bases, it is then straightforward to determine whether the values of the exponents are different, since associated to typical and average quantities, respectively. Indeed, it is immediately obvious from the figure that the data for $\mathcal{T}_c(\mathcal{N})$ and $\delta\mathcal{T}_c(\mathcal{N})$ (as the ones for $\mathcal{T}_c'(\mathcal{N})$ and $\delta\mathcal{T}_c'(\mathcal{N})$) display essentially the same $\mathcal{N}$-dependence. This result implies that the pseudo first order transition scenario cannot describe the observed behaviour in the present case, given that the mean value and the fluctuations of the pseudo-$\mathcal{T}_c$ are ruled by the same exponent ν_r (that appears to be larger than 2).

For further deepened analysis, [Tab. 2.3] reports evaluations of $\mathcal{T}_c$ and $1/\nu_{r,1}$ (from $\mathcal{T}_c(\mathcal{N})$), and of $1/\nu_{r,2}$ (from $\delta\mathcal{T}_c(\mathcal{N})$), obtained disregarding the chains of length $\mathcal{N} < \mathcal{N}_0$, for different $\mathcal{N}_0$ values. This analysis does not reveal any evident finite size corrections to scaling in the behaviour of $\mathcal{T}_c(\mathcal{N})$. On the other hand, $1/\nu_{r,2}$ appears to display a weak dependence on $\mathcal{N}_0$. In detail, it decreases to the value $1/\nu_{r,2} \sim 0.33 - 0.34$ when only chains of length $\mathcal{N} \geq 1000$ are considered. Thus we are led to similar values for the two exponents, with $1/\nu_{r,1} = 1/\nu_{r,2} = 1/\nu_r = 0.35 \pm 0.05$, and accordingly the evaluations: $\nu_r = 2.9 \pm 0.04$, and $c_r = 1 + 1/\nu_r = 1.35 \pm 0.05$. Furthermore, the analysis of the mean value and of the fluctuations of $\mathcal{T}_c'(\{\epsilon_i\},\mathcal{N})$ leads to evaluations ($1/\nu_{r,1} = 0.44 \pm 0.06$ and $1/\nu_{r,2} = 0.38 \pm 0.01$) that are consistent with the previous ones, even though in this case only chains of length $\mathcal{N} \leq 2500$ are considered.

Pseudo-critical temperatures appear thus to represent particularly interesting observables for the model here, with their scaling behaviour clearly in accordance with the usual scenario corresponding to relevance of disorder, the same as the one in the analysis of random ferromagnets [19, 62, 63]: the numerical analysis provides evidence for a smooth phase transition (as already found in [37]), the thermodynamic limit behaviour being ruled by a single correlation length, in agreement notably with the mathematical findings in [20, 21]. More precisely, our present best evaluation of the random

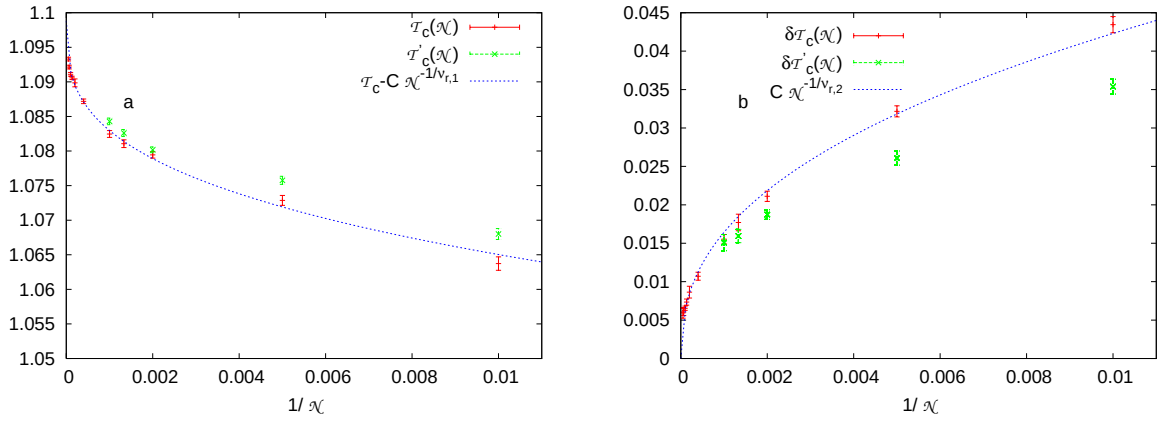

Figure 2.8: **a) Plots of the mean values, $\mathcal{T}_c(\mathcal{N})$ and $\mathcal{T}_c'(\mathcal{N})$, of the pseudo-critical temperatures as functions of $1/\mathcal{N}$.** The $\mathcal{T}_c(\mathcal{N})$ values are plotted for the various chain lengths (red marks with error bars), corresponding for a given $\mathcal{N}$ to the mean value of $\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})$, associated with the temperature at which the susceptibility reaches its global maximum, for the various sequences. In addition, mean values of $\mathcal{T}_c'(\{\epsilon_i\},\mathcal{N})$ are plotted for $\mathcal{N} \leq 2500$ (green marks with error bars), defined as the crossing points of the order parameters $\Theta(\{\epsilon_i\}, t\mathcal{N}, \mathcal{T})$, for $t = 1, 2, 4$. The dotted line corresponds to the best fit of $\mathcal{T}_c(\mathcal{N})$, according to the scaling law $\mathcal{T}_c(\mathcal{N}) \simeq \mathcal{T}_c + C\mathcal{N}^{-1/\nu_{r,1}}$ (with $\mathcal{T}_c = 1.098 \pm 0.001$ and $1/\nu_{r,1} = 0.33 \pm 0.03$). **b) Plots of the fluctuations $\delta\mathcal{T}_c(\mathcal{N})$ and $\delta\mathcal{T}_c'(\mathcal{N})$, associated with $\mathcal{T}_c(\mathcal{N})$ and $\mathcal{T}_c'(\mathcal{N})$ in a), as functions of $1/\mathcal{N}$.** The dotted line corresponds to the best fit for $\delta\mathcal{T}_c(\mathcal{N})$, according to the scaling law $\delta\mathcal{T}_c(\mathcal{N}) \propto \mathcal{N}^{-1/\nu_{r,2}}$ (with $1/\nu_{r,2} = 0.405 \pm 0.01$).

$\mathcal{N}_0$	$\mathcal{T}_c$	$1/\nu_{r,1}$	$1/\nu_{r,2}$
100	1.098±0.001	0.33 ± 0.03	0.405 ± 0.01
200	1.101±0.002	0.27 ± 0.04	0.38 ± 0.01
500	1.101±0.003	0.27 ± 0.06	0.37 ± 0.01
750	1.099±0.003	0.32 ± 0.08	0.365 ± 0.01
1000	—	—	0.34 ± 0.02
2500	—	—	0.33 ± 0.06

Table 2.3: Evaluations of $\mathcal{T}_c$, $1/\nu_{r,1}$ and $1/\nu_{r,2}$ obtained disregarding chains of length $\mathcal{N} < \mathcal{N}_0$, for the different $\mathcal{N}_0$ values considered. For $\mathcal{T}_c$ and $1/\nu_{r,1}$ evaluations are from $\mathcal{T}_c(\mathcal{N})$ data, and for $1/\nu_{r,2}$ from $\delta\mathcal{T}_c(\mathcal{N})$ data (see the text for details).

critical point correlation length exponent ($\nu_r = 2.9 \pm 0.04$) is perfectly consistent with the previous evaluation in [37], obtained through the scaling of the maximum of the specific heat averaged over disorder in the standard way. Nonetheless, such an evaluation is both more refined and more reliable than the previous one, too.

Importantly, it is also clear that, for the pseudo-critical temperature oriented analyses, asymptotic behaviours appear to be reached for chain lengths shorter than those considered in the present study (up to $\mathcal{N} = 2 \cdot 10^4$), despite the strong finite size corrections to scaling characterizing the behaviour of the various thermodynamical observables in this disordered model [36, 37]. Even though, from this point of view, the pseudo-$\mathcal{T}_c$ appears among the most interesting observables to look at, the overall situation that concerns the finite size effects is anyway expected to be related to the choice of parameters. As we are going to discuss, in order to reach a quantitative description of the dependency of the behaviour of the model on these parameters it would be relevant to determine a crossover chain length $\mathcal{N}^*$, below which also the present data could agree with a pseudo first order transition.

With this respect, in order to grasp the behaviour of the model in the thermodynamic limit, looking in detail at [Fig. 2.8], it appears important, on qualitative bases, to have $\mathcal{N} > 1000$ (*i.e.* $1/\mathcal{N} < 0.001$). Indeed, with $\mathcal{N} \leq 1000$, the average $\mathcal{T}_c(\{\epsilon_i\}, \mathcal{N})$, as a function of $1/\mathcal{N}$, would appear instead to be adequately fitted by a straight line, while this is not true for $\delta\mathcal{T}_c(\{\epsilon_i\}, \mathcal{N})$. Thus the behaviour of these observables would be in agreement with a transition with $\nu_{r,2} > \nu_{r,1} = \nu_p = 1$. Accordingly, $\mathcal{N}^* \sim 1000$ can be retained as the evaluation of a crossover chain length suggested by data on the mean value and on the fluctuations of the pseudo-$\mathcal{T}_c$ in our case. Clearly, this value corresponds to an evaluation *from below*, just as the considered observables appear to be the less affected by corrections to scaling.

2.7.4 On the behaviour of the order parameter and of the susceptibility

The behaviour of $\overline{\overline{\Theta(\{\epsilon_i\}, \mathcal{T}, \mathcal{N})}}$ is plotted in [Fig. 2.9], following the two definitions of pseudo-critical temperatures that we considered. Indeed, the present analysis aims to understand at a qualitative level the general results on quantities averaged by taking into account the pseudo-$\mathcal{T}_c$. We recall that the expected scaling law of this observable is $\mathcal{N}^{1-1/\nu_r}\tilde{\Theta}[(\mathcal{T}_c - \mathcal{T})\mathcal{N}^{1/\nu_r}]$, with $\tilde{\Theta}$ a scaling function, that can be derived from (2.5) by using $e_r = \beta_r = \nu_r - 1$. However, we do not attempt to perform any finite size scaling analysis for the order parameter since, independently of the averaging method, the behaviour of this observable is not in agreement with such a scaling law. This feature was observed both in [38] and in our previous work [37], too.

From a qualitative point of view, the data displayed in [Fig. 2.9] show unambiguously that, with varying $\mathcal{N}$, the order parameters do not cross at the same temperature. This situation stands in sharp contrast with the one characterizing the pure model and also with the results reported in [38, 39]. As a matter of fact, this feature, clearly in disagreement with the possibility of a pseudo first order character for the behaviour of the model, was already observed in [37]. With the present analysis, it further appears that this result does not depend on the used averaging method. In fact, this result concerning crossovers is more evident for shorter chain lengths, but it can be observed also for the largest considered ones. Finally, such a result appears to hold independently of the definition of the pseudo-critical temperature, though the positions of the crossing points vary still more rapidly with $\mathcal{N}$ when the average is performed by using the definition of the pseudo-$\mathcal{T}_c$ that we proposed.

Importantly, the analysis in terms of pseudo-critical temperatures reduces, in general, the importance of the finite size effects, hence allowing notably to extrapolate the correct thermodynamic limit behaviour from shorter chain lengths (*i.e.*, decreasing the effective crossover length, $\mathcal{N}^*$, value). In order to make evident this point, we consider in detail the behaviour of the average susceptibility, whose evaluation is expected to be the most sensitive to the different ways of averaging. Indeed, notably in the context of the study here, that resorts extensively to the definition of the pseudo-critical temperature as the position of the global maximum of the susceptibility for a given sequence,

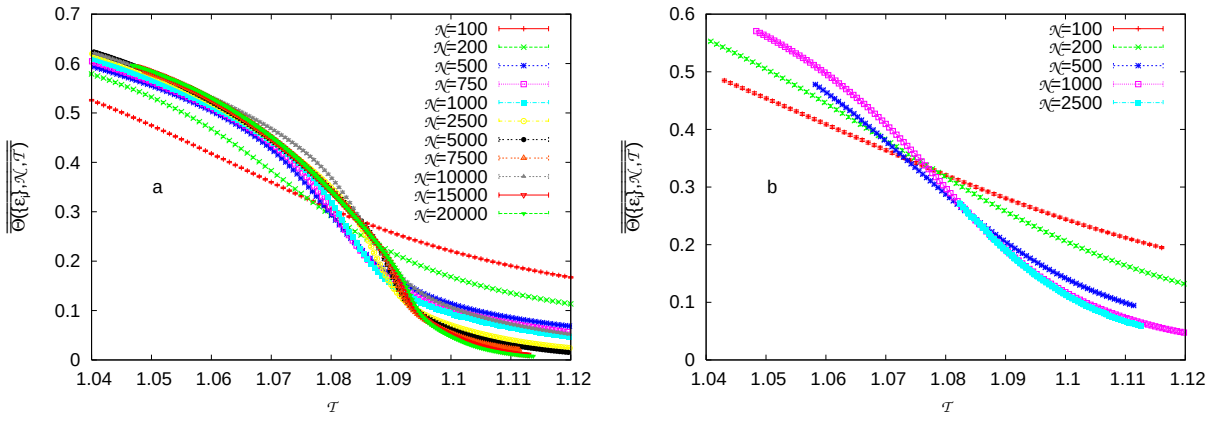

Figure 2.9: **Plots of $\overline{\overline{\Theta(\{\epsilon_i\},\mathcal{N},\mathcal{T})}}$**, with the average performed by taking into account the pseudo-critical temperature, following the two considered definitions: a) the pseudo-critical temperature, $\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})$, is defined as the value at which the global maximum of the susceptibility is reached; b) for chains of length $\mathcal{N} \leq 2500$, the pseudo-critical temperature, $\mathcal{T}_c'(\{\epsilon_i\},\mathcal{N})$, is defined as the crossing point of the plots of $\Theta(\{\epsilon_i\},t\mathcal{N},\mathcal{T})$ (for $t = 1, 2$ and 4, obtained from the concatenation of t copies of the original sequence).

the quantity $\overline{\overline{\chi(\{\epsilon_i\},\mathcal{T},\mathcal{N})}}$ is expected to best highlight, with its possible divergence, a pseudo first character in the behaviour of the model (see (2.21)). Furthermore, the maximum of this quantity is expected to behave as a typical quantity, in the sense of being the observable the less affected by fluctuations in the pseudo-$\mathcal{T}_c$ itself (see again (2.21)). From this point of view the study of this observable is expected to best highlight possible differences between typical and average behaviours, too.

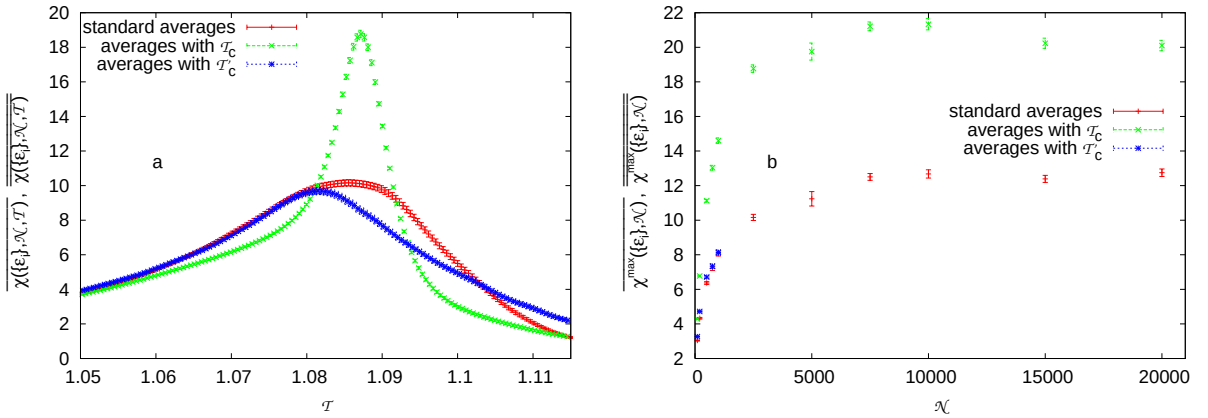

Figure 2.10: **a) Plots of $\overline{\chi(\{\epsilon_i\},\mathcal{T},\mathcal{N})}$ and of $\overline{\overline{\chi(\{\epsilon_i\},\mathcal{T},\mathcal{N})}}$ for chain length $\mathcal{N} = 2500$.** Here, the average is performed both in the standard way (these are the same data as in [37]) and by taking into account the pseudo-critical temperature. In the last case, we follow the two considered definitions (*i.e.*, with $\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})$ the temperature at which, for the given sequence, the susceptibility reaches its global maximum and with $\mathcal{T}_c'(\{\epsilon_i\},\mathcal{N})$ the temperature corresponding to the crossing point of the order parameter, for the given sequence, with those of the sequences obtained by the concatenation procedure). **b) Plots of $\overline{\chi^{max}(\mathcal{N})}$ and of $\overline{\overline{\chi^{max}(\{\epsilon_i\},\mathcal{T},\mathcal{N})}}$.** Here, too, the average is performed both in the standard way (these are the same data as in [37]) and by taking into account the pseudo-critical temperature. In the last case, we follow the two considered definitions.

In detail, the results on the susceptibility data, for $\mathcal{N} = 2500$, are plotted in [Fig. 2.10a] for $\overline{\chi(\{\epsilon_i\},\mathcal{T},\mathcal{N})}$ (as obtained from (2.6) and corresponding to the same data as in [37]) and for $\overline{\overline{\chi(\{\epsilon_i\},\mathcal{T},\mathcal{N})}}$ (as obtained from (2.12)), using the two definitions of the pseudo-critical temperature

($\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})$ and $\mathcal{T}'_c(\{\epsilon_i\},\mathcal{N})$, respectively). As expected, the figure shows that if the behaviour of the model would have had a pseudo first order character, it should have been evident from our data on $\overline{\overline{\chi}}$ evaluated by defining the pseudo-$\mathcal{T}_c$ as the abscissa of the global maximum of the susceptibility (for a given sequence), *i.e.*, $\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})$. In this case, the average susceptibility reaches a definitely highest maximum. The observation thereby confirms that the considered $\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})$ is particularly appropriate both for assessing disorder relevance and for determining the behaviour of the model with respect to the usual scenario.

In fact, when looking at the scaling behaviours of the maxima values as functions of $\mathcal{N}$ in [Fig. 2.10b], one finds a qualitative good agreement between the different considered ways of performing the averages. In particular, in all cases the maximum of the average susceptibility displays a crossing to an $\mathcal{N}$-independent regime for the largest considered chain lengths, clearly showing that there is no difference in the behaviour of typical and average quantities. The data therefore further support a smooth transition (with $\gamma_r = \alpha_r \leq 0$, and accordingly $\nu_r = 2-\alpha_r \geq 2$). From a different point of view, these findings also clearly confirm qualitatively that the thermodynamic limit behaviour of the model is described by a single correlation length, in agreement with the result following which the mean value and the fluctuations of the pseudo-$\mathcal{T}_c$ scale with the same exponent (namely, $\nu_r = 2.9 \pm 0.4$).

It is worth recalling that in our previous work [37], by fitting the data for the maximum of the specific heat (averaged in the standard way) to the law $C_1 - C_2\mathcal{N}^{\alpha_r/\nu_r}$, we obtained (with the first two points disregarded) $\alpha_r/\nu_r = -0.3 \pm 0.1$, and hence $\nu_r = 2.9 \pm 0.6$. Thus, it would be meaningless to repeat the analysis on $\overline{\overline{\chi}}^{max}$ here.

Inasmuch as the data for $\overline{\max_{\mathcal{T}}\{\chi(\{\epsilon_i\},\mathcal{T},\mathcal{N})\}}$ display an abrupt change as function of $\mathcal{N}$, it is evidently appropriate to introduce a crossover chain length $\mathcal{N}^*$, for characterizing the slow approach of the model to the asymptotic regime. More precisely, we observe a shift from a short chain increasing behaviour to a long chain nearly constant one, around a value of $\mathcal{N}^* \sim 2500$, that is compatible both with the evaluation from below ($\mathcal{N}^* \sim 1000$) obtained qualitatively from the scaling of the mean value and of the fluctuations of the pseudo-critical temperature and with the evaluation from above ($\mathcal{N}^* \sim 2.500 - 5.000$), that is the most accurate evaluation of this quantity that we obtain according to the behaviour of the loop-length probability distribution, as we show in the following Section 2.7.5.

In order to conclude the discussion on the behaviour of the susceptibility, let us show the data on the non self-averageness parameter, $\mathcal{R}$, as defined in (2.13)), at least in this case, in which, on the basis of the observed single sequence multi-peak behaviour, the strong non self-averaging behaviour is expected to be particularly evident, though this observable is singular but does not diverge in the random case.

As already recalled, both in the usual case of disorder relevance and within the pseudo first order transition scenario, one expects strong non self-averageness in the thermodynamic observables that are singular at the critical point. Accordingly, the parameter $\mathcal{R}$, that measures the relative fluctuations of the observable averaged over disorder in the standard way, should display a constant behaviour instead of decreasing as a function of the chain length $\mathcal{N}$ at $\mathcal{T}_c$.

In fact, numerical evidence was reported in [39] for such a behaviour of the non self-averageness parameter related to $\overline{\Theta}$, for the disordered PS models considered with different c_p values, and notably for the one with $c_p = 2.15$. It is therefore reasonable to assume that the same result should hold in the present system. Indeed, as we already noticed, since the single sequence order parameter displays multi-step behaviour in a significant fraction of the samples, the sequence-to-sequence fluctuations of this observable, averaged in the standard way, should play a role at least of the same importance as in the model studied with $c_p = 2.15$ in [39], in which this behaviour is not observed.

In this background, we instead focus here on $\mathcal{R}_\chi(\mathcal{N},\mathcal{T})$, whose behaviour as function of $\mathcal{T}$ is plotted in [Fig. 2.11], for the different considered chain lengths. First, letting aside the two shortest lengths $\mathcal{N} = 100$ and $\mathcal{N} = 200$, in the plots of $\mathcal{R}_\chi(\mathcal{N},\mathcal{T})$ no evident dependence on the chain length $\mathcal{N}$ is observed for the heights of peaks, thus confirming the expected strong non self-averageness of this observable in the present model. The plots in [Fig. 2.11] display rather irregular behaviours in

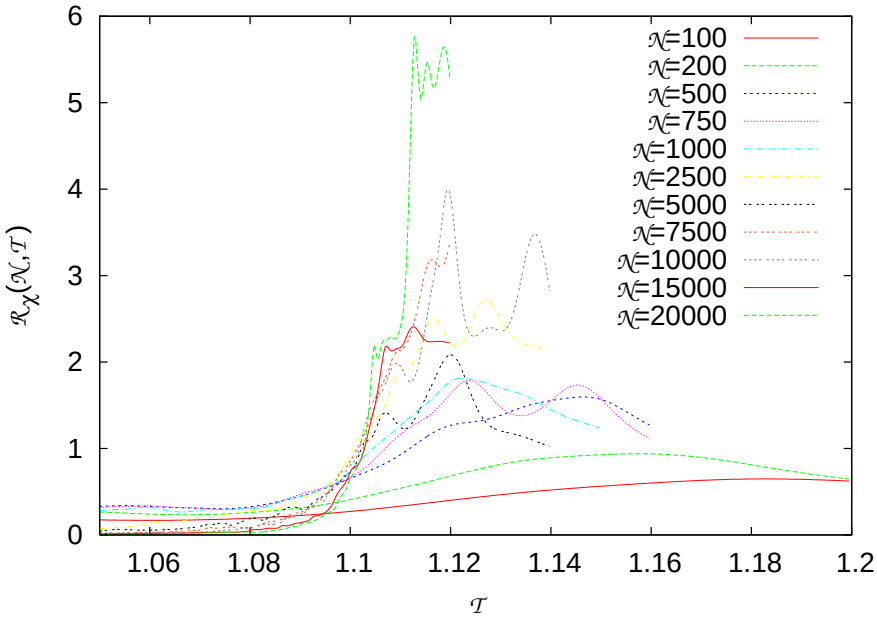

Figure 2.11: **Plots of the non self-averaging parameter related to the susceptibility, $\mathcal{R}_\chi(\mathcal{N},\mathcal{T})$, as function of the temperature, for the different considered chain lengths.** No evaluations of the errors are performed, even though they are expected to be at most of the order of the observed oscillations of the quantity.

the whole high-$\mathcal{T}$ region, even though it is in general expected that $\mathcal{R} \sim 1/\mathcal{N}$ both above and below the critical point. This observation can be explained by the fact that here the disorder couples only to the low-temperature phase, with the order parameter being zero in the thermodynamic limit for $\mathcal{T} > \mathcal{T}_c$, where the two DNA strands are only linked at the origin. It is then particularly difficult to evaluate correctly in this region the fluctuations due to disorder of the average susceptibility.

Nevertheless, it is interesting to underline in [Fig. 2.11], that the plots for different $\mathcal{N}$-values display similar behaviours near the thermodynamic limit critical temperature $\mathcal{T}_c \simeq 1.1$. In particular, a steep increase is observed in the plots immediately above $\mathcal{T}_c \simeq 1.1$, with the steepness increasing with the chain length, and for temperatures approaching $\mathcal{T}_c$ from above for increasing chain lengths. In more quantitative terms, for the different considered $\mathcal{N}$-values, the position of the highest $\mathcal{R}_\chi(\mathcal{N},\mathcal{T})$ peak, corresponding to the best evaluation of the abscissa of its global maximum as allowed by our present statistics, appears to be interpretable as the mean value of a quantity behaving as a pseudo critical temperature (or in any event as an $\mathcal{N}$-depending evaluation of the thermodynamic limit critical temperature $\mathcal{T}_c$), that we denote $\mathcal{T}_c^{\mathcal{R}}(\mathcal{N})$. In fact, we checked that the scaling law $\mathcal{T}_c^{\mathcal{R}}(\mathcal{N}) = \mathcal{T}_c + C\mathcal{N}^{-1/\nu_r}$ is valid for this observable, too. With respect to the previous cases, corresponding to the two different definitions of $\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})$, it is clear that the constant has the opposite sign, with anyway the fit leading to compatible estimations of $\mathcal{T}_c$ and ν_r.

2.7.5 A detailed analysis on the distribution probability of the loop length

It was further hypothesized in [38, 39] that the presence of two correlation lengths ruled by different critical exponents could be inferred from the probability distribution of the loop-lengths. Qualitatively, one would expect, in particular, different behaviours for $\overline{\log P(\{\epsilon_i\},\mathcal{N},\mathcal{T})}$, that is obtained by performing the (standard) average after taking the logarithm, and for $\log \overline{P(\{\epsilon_i\},\mathcal{N},\mathcal{T})}$, that is obtained by taking the logarithm after performing the (standard) average. We notice that such a conclusion did not seem to be confirmed by the results in [37], in which the data on these two quantities at $\mathcal{T} = \mathcal{T}_c$ did not display asymptotically detectable differences. Moreover, apart from

the mathematical results in [20, 21], the present analysis on the behaviour of the pseudo-critical temperature and the maximum of the non standardly averaged susceptibility definitely rule out the possibility of a pseudo first order transition. Nonetheless, it appears desirable to complete the present investigation by analysing this point in detail, too, with a careful study of the data on the probability distribution of the loop length on the whole temperature range. Above all, as we are going to discuss, such an analysis allows us to give both a clear definition and a refined evaluation of the crossover chain length $\mathcal{N}^*$. In fact, this quantity plays a key role in our semi-quantitative explanation of the apparent contradiction between the different numerical results, that is the starting point for our subsequent study that aims to formulate such an explanation in more rigorous terms.

The quantity:

$$P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l) \equiv (2l)^{c_p} P(\{\epsilon_i\},\mathcal{N},\mathcal{T},l), \tag{2.22}$$

was introduced in our previous work [37], as more appropriate than the loop probability distribution itself for capturing relevance of disorder. Indeed, at the critical point, the logarithm of $P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)$ is expected to be not constant and proportional to $(c_p - c_r)\log l + C$ as soon as $c_r < c_p$.

In more detail, we consider both $\log \overline{P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)}$ and $\overline{\log P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)}$, that, as just recalled, should allow to capture the behaviours of the average and typical correlation lengths respectively (with in fact the second case better described as a mixed average).

The behaviour of these quantities for $\mathcal{T} \simeq \mathcal{T}_c$ was found compatible with a smooth transition in [37], but it is noticeable that the evaluated critical exponent c_r depends on $\mathcal{N}$. Therefore, in that work, one could only give an upper limit on c_r, *i.e.*, $c_r \leq 1.5$, whose correctness is confirmed by our present evaluation $c_r = 1.35 \pm 0.05$ obtained from the pseudo-$\mathcal{T}_c$ behaviour. In fact, this $\mathcal{N}$-dependency is even more evident when considering the average quantity $\overline{\log P'(\{\epsilon_i\},\mathcal{N},\mathcal{T}_c,l)}$, that displays more important deviations from the expected behaviour ($\propto \log l$) than $\log \overline{P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)}$.

In order to further clarify the situation, and particularly with the aim of better characterizing the finite size behaviour of the present model, we are led to study in detail both of these quantities on the whole relevant $\mathcal{T}$-range, by introducing effective ($\mathcal{N}$-dependent) critical exponents $c_{r,1}(\mathcal{N})$ and $c_{r,2}(\mathcal{N})$, as well as correlation lengths $\xi_1(\mathcal{N},\mathcal{T})$ and $\xi_2(\mathcal{N},\mathcal{T})$.

We found that the data concerning $\log \overline{P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)}$ are accurately described by the expected scaling law, which can be easily obtained from (2.19). Accordingly, the behaviour of the observable, which should be ruled by the fluctuations of the pseudo-$\mathcal{T}_c$, is in agreement with:

$$\log \overline{P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)} \simeq (c_p - c_{r,2}(\mathcal{N}))\log l - l/\xi_2(\mathcal{N},\mathcal{T}) + C(\mathcal{N}). \tag{2.23}$$

On the other hand, in order to obtain a satisfactory fit (within the errors) of the data concerning $\overline{\log P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)}$ (with such an analysis expected to better capture the typical loop-length probability distribution behaviour), it appears necessary to introduce, in addition, a quadratic contribution in l, *i.e.*:

$$\overline{\log P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)} \simeq (c_p - c_{r,1}(\mathcal{N}))\log l - l/\xi_1(\mathcal{N},\mathcal{T}) - C_1(\mathcal{N})l^2 + C_2(\mathcal{N}), \tag{2.24}$$

with the constant $C_1(\mathcal{N})$ tending to zero roughly proportionally to $1/\mathcal{N}$.

Such an approach appears adequate to describe the strong finite size corrections to scaling that characterizes the present model, and in the meanwhile it allows a clear definition of $\mathcal{N}^*$. In fact, the crossover between a short chain length regime, in which the effective exponents would be in agreement with a pseudo first order transition, and a long chain length one, in which they would be compatible with the asymptotic values, turns out to be quite abrupt. Therefore, one can obtain a quantitative evaluation of the crossover chain length (here, an evaluation *from above*), as the length $\mathcal{N}^*$ beyond which $c_{r,1}(\mathcal{N}) \sim c_{r,2}(\mathcal{N}) \sim c_r \simeq 1.35$ and $\lim_{\mathcal{T}\to\mathcal{T}_c^-} \xi_1(\mathcal{N},\mathcal{T}) \sim \lim_{\mathcal{T}\to\mathcal{T}_c^-} \xi_2(\mathcal{N},\mathcal{T}) \sim (\mathcal{T}_c - \mathcal{T})^{-\nu_r}$, with $1/\nu_r = c_r - 1 \simeq 0.35$.

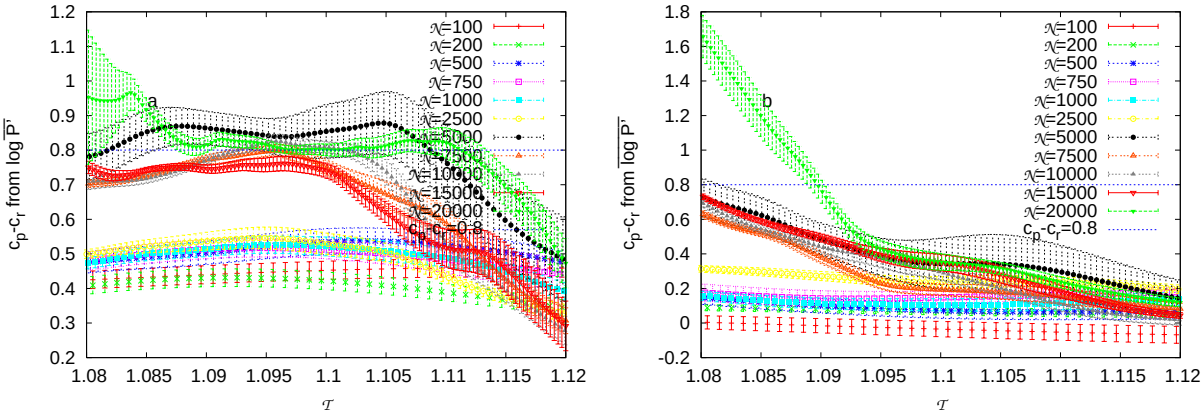

Figure 2.12: **a) Evaluation, for the various considered chain lengths, of $c_p - c_{r,2}(\mathcal{N})$.** Here the evaluations are obtained from the fits (three parameters) of $\log \overline{P'(\{\epsilon_i\}, \mathcal{N}, \mathcal{T}, l)}$ to (2.23), as function of the temperature. **b) Evaluation, for the various considered chain lengths, of $c_p - c_{r,1}(\mathcal{N})$.** Here the evaluations are obtained from the fits (four parameters) of $\overline{\log P'(\{\epsilon_i\}, \mathcal{N}, \mathcal{T}, l)}$ to (2.24), as function of the temperature. In both the analyses the errors are only indicative, since the results display some dependence on the l-range (we show the results obtained by taking the range $l \in [3, \mathcal{N}/3]$, that corresponds to a reasonable choice), in particular for the shortest chain lengths. For comparisons, the expected asymptotic behaviour ($c_p - c_r \simeq 0.8$) is also represented in both panels.

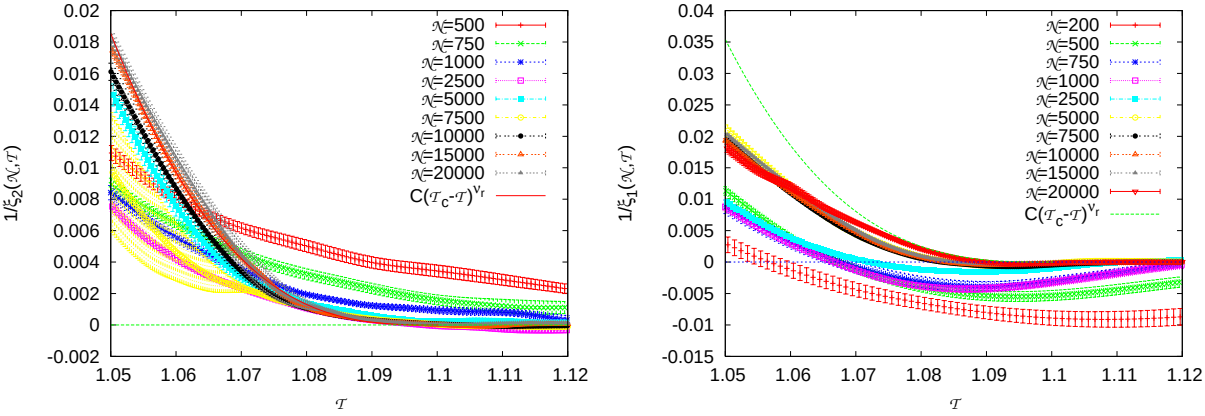

Figure 2.13: **a) Evaluation of the inverse of the correlation length as function of the temperature for the various chain lengths considered.** Here the evaluations are obtained from the fits (three parameters) of $\log \overline{P'(\{\epsilon_i\}, \mathcal{N}, \mathcal{T}, l)}$ to (2.23). **b) Evaluation of the inverse of the correlation length $\xi_1(\mathcal{N}, \mathcal{T})$ as function of the temperature for the various chain lengths considered.** Here the evaluations are obtained from the fits (four parameters) of $\overline{\log P'(\{\epsilon_i\}, \mathcal{N}, \mathcal{T}, l)}$ to (2.24). The same remarks as in [Fig. 2.12] apply to the significance of errors. For comparisons, the expected asymptotic behaviour ($1/\xi(\mathcal{T}) \propto (\mathcal{T}_c - \mathcal{T})^{2.9}$) is also represented in both panels.

The detailed results for the loop-length probability distribution exponents and the inverse of the correlation lengths are presented in [Fig. 2.12] and [Fig. 2.13], respectively, associated with (2.23) and (2.24) in each of the two cases. More precisely, the figures plot the evaluations of $c_p - c_{r,2}(\mathcal{N})$, $c_p - c_{r,1}(\mathcal{N})$, $1/\xi_1(\mathcal{N},\mathcal{T})$ and $1/\xi_2(\mathcal{N},\mathcal{T})$, as obtained from the fits above, as functions of the temperature for the different considered chain lengths. These plots moreover allow to compare the results with $c_p - c_r \simeq 0.8$ and $1/\xi(\mathcal{T}) \sim (\mathcal{T}_c - \mathcal{T})^{2.9}$ for $\mathcal{T} \to \mathcal{T}_c^-$ (the asymptotic inverse of the correlation length being zero in the whole high temperature phase).

The plots in [Fig. 2.12] and [Fig. 2.13] clearly illustrate the strong $\mathcal{N}$-dependence of the considered quantities. Nonetheless, as shown in particular in [Fig. 2.12a], for chains of length $\mathcal{N} \geq 5000$, by allowing for a non-zero effective $1/\xi_2(\mathcal{N},\mathcal{T})$, the evaluations of $c_p - c_{r,2}(\mathcal{N})$ from $\log \overline{P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)}$ are consistent with $c_r \simeq 1.35$ over a large temperature range around $\mathcal{T}_c$. Accordingly, for the disordered PS model with $c_p = 2.15$ studied, it appears that $\mathcal{N}^* \sim 2500 - 5000$ can be taken as our best evaluation from above of the crossover length $\mathcal{N}^*$. This conclusion holds for the analysis of $\log \overline{P'(\{\epsilon_i\},\mathcal{N},\mathcal{T},l)}$ data, too, even though more significant corrections are involved in this case.

It is also very illustrative to outline that, by limiting the analysis to chains of length $\mathcal{N} < 2500$ one could reach interpretations significantly different from those obtained within the long chain length regime: one would get from [Fig. 2.12b] an evaluation for $c_p - c_{r,1}(\mathcal{N})$ essentially compatible with the zero value characteristic of the pure case (*i.e.* $c_{r,1} = c_p = 2.15$ that would imply $\nu_{r,1} = \nu_p = 1$), as well as (from [Fig. 2.12a]) an evaluation for $c_{r,2}(\mathcal{N})$ not significantly larger than 1.5 (*i.e.*, $\nu_{r,2} \sim 2$, since $c_{r,2} = 1 + 1/\nu_{r,2}$). With this respect, the analysis here confirms that, in the regime below the crossover chain length $\mathcal{N}^*$, it is difficult to extrapolate the correct thermodynamic limit behaviour. Indeed, in this regime, the results turn out to be explainable according to the picture that implies a pseudo first order phase transition, as proposed in [38, 39].

The importance of finite size effects is further highlighted in [Fig. 2.13]. This figure allows to compare the behaviour of $1/\xi_2(\mathcal{N},\mathcal{T})$ ([Fig. 2.13a], as obtained from (2.23)), with that of $1/\xi_1(\mathcal{N},\mathcal{T})$ ([Fig. 2.13b], as obtained from (2.24)), with the plots clearly showing that the two quantities approach the asymptotic limit $\propto (\mathcal{T}_c - \mathcal{T})^{\nu_r}$ with $\nu_r \simeq 2.9$ from opposite sides. In particular, for chains of short lengths $\mathcal{N} < \mathcal{N}^*$, it appears that the value of $1/\xi_2(\mathcal{T},\mathcal{N})$ is definitely different from zero, on the whole $\mathcal{T}$-range considered. It can be also observed that, in the same short chain length regime, when looking at the averages performed after taking the logarithm, one obtains negative (physically meaningless) values of $\xi_1(\mathcal{T},\mathcal{N})$ on a large part of the $\mathcal{T}$-range (however, we remind that in this case it appears necessary to insert also a quadratic term in l in order to perform the data fit). Even though the here presented data do not allow a more quantitative analysis, these results further imply that below the crossover (*i.e.*, for chain of lengths $\mathcal{N} < 2500$) one gets numerical evidence for the presence of two correlation lengths (and accordingly for two different critical exponents $\nu_{r,2} \neq \nu_{r,1}$).

On the other hand, the behaviours of ξ_1 and ξ_2 for $\mathcal{N} > \mathcal{N}^*$ are in agreement over a large $\mathcal{T}$-range, within the error margins, with a correlation length exponent $\nu_r \simeq 2.9$, as clearly shown in [Fig. 2.14], that displays the plots of $1/\xi_2(\mathcal{N},\mathcal{T})$ and $1/\xi_1(\mathcal{N},\mathcal{T})$ for the three considered largest chain lengths. The fact that the two quantities approach the same behaviour (*i.e.*, the asymptotic one that is predictable on the basis of our evaluations of ν_r and $\mathcal{T}_c$) is particularly meaningful: these observables are very demanding to be measured, and moreover they are the ones expected to be the most affected by corrections to scaling. Therefore, the result confirms that for $\mathcal{N} \sim 10^4$ the asymptotic regime of the studied model is definitely reached.

Finally, for the different chain lengths, [Fig. 2.15] displays plots of $\log \overline{\overline{P'(\{\epsilon_i\},\mathcal{N},\mathcal{T}_c,l)}}$ and $\log \overline{\overline{P'(\{\epsilon_i\},\mathcal{N},\mathcal{T}_c,l)}}$ with the averages that take into account the pseudo-critical temperature, following (2.12). More precisely, in this case, for any given sequence, the loop-length probability distribution which contributes to the average is the one evaluated at $\mathcal{T}_c(\{\epsilon_i\},\mathcal{N})$, where the susceptibility of the sequence reaches its global maximum. If comparing these plots to those of $\log \overline{P'(\{\epsilon_i\},\mathcal{N},\mathcal{T}_c,l)}$ and $\log \overline{P'(\{\epsilon_i\},\mathcal{N},\mathcal{T}_c,l)}$ that were presented in [37], it becomes once more clear that the analysis in terms of pseudo-critical temperatures allows to reduce the importance of finite size corrections to scaling.

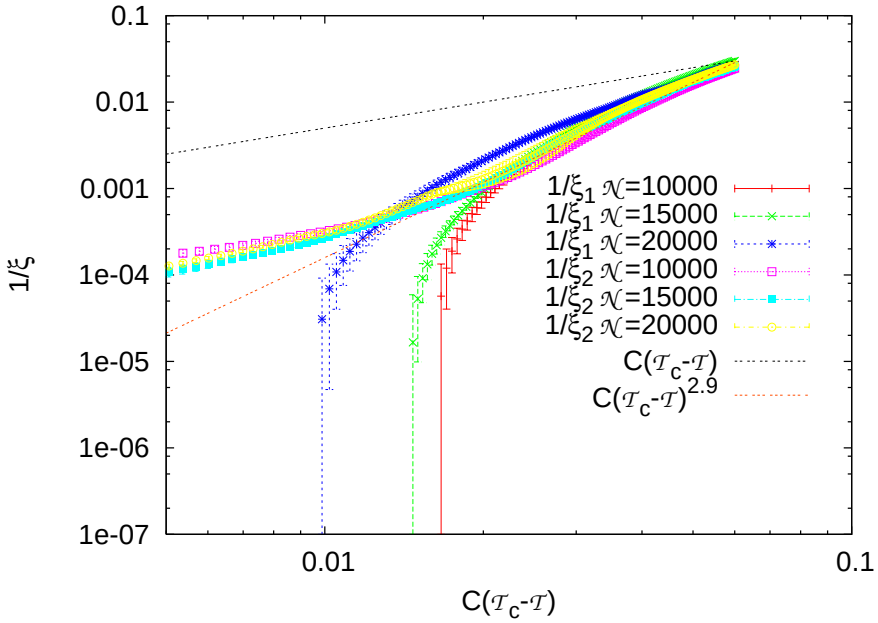

Figure 2.14: **Plots of $1/\xi_2(\mathcal{N},\mathcal{T})$ and $1/\xi_1(\mathcal{N},\mathcal{T})$.** The quantities are plotted on log-log scale as functions of $(\mathcal{T}_c - \mathcal{T})$ (with $\mathcal{T}_c \simeq 1.1$) for the three longest chain lengths considered ($\mathcal{N} = 10000$, $\mathcal{N} = 15000$ and $\mathcal{N} = 20000$; same data as in [Fig. 2.13]). Data are plotted against the expected asymptotic behaviour of the inverse correlation length $1/\xi(\mathcal{T}) \propto (\mathcal{T}_c - \mathcal{T})^{\nu_r}$, with $\mathcal{T}_c = 1.1$ and $\nu_r = 2.9$. For comparisons, the behaviour of $1/\xi(\mathcal{T}) \propto (\mathcal{T}_c - \mathcal{T})$ is also displayed.

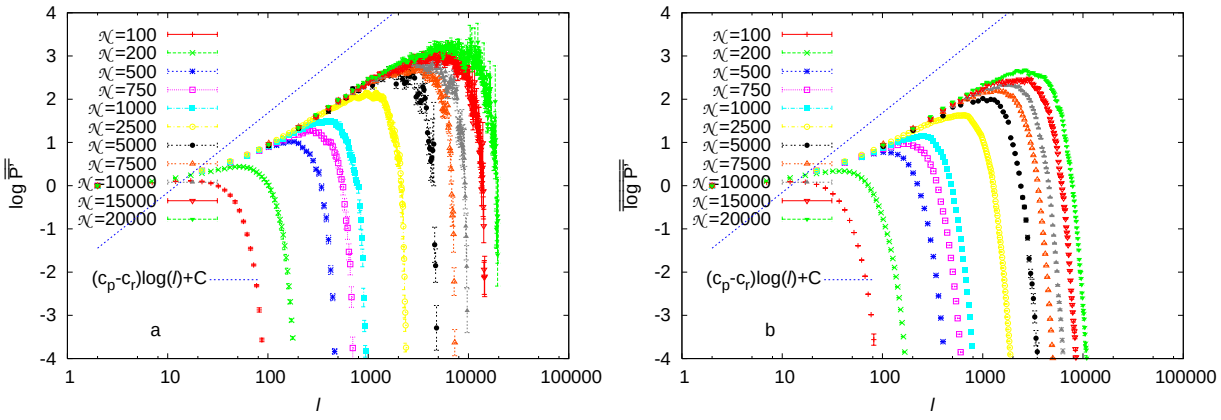

Figure 2.15: **a) Plots of** $\overline{\overline{\log P'(\{\epsilon_i\},\mathcal{N},\mathcal{T}_c,l)}}$**.** The quantity is plotted as functions of l at $\mathcal{T} = \mathcal{T}_c \simeq 1.1$, for the different considered chain lengths. **b) Plots of** $\overline{\overline{\log P'(\{\epsilon_i\},\mathcal{N},\mathcal{T}_c,l)}}$**.** The quantity is plotted as functions of l at $\mathcal{T} = \mathcal{T}_c \simeq 1.1$, for the different considered chain lengths. For both a) and b) the averages are performed by taking into account the pseudo-critical temperature, following (2.12). All curves are shifted arbitrarily, setting them to zero for $l = 2$. The expected asymptotic limit behaviour $\propto (c_p - c_r)\log l + C$, with $c_p - c_r = 0.8$ (*i.e.*, $c_r = 1.35$), is also plotted, in both panels, as a dotted line.

Here we limit ourselves to observe that, first of all, the present data that correspond to the average performed after taking the logarithm display a behaviour more similar to the one in which the logarithm is taken after the average. Moreover, both quantities approach qualitatively more rapidly the asymptotic behaviour $\propto (c_p - c_r)\log l + C$, with $c_r \simeq 1.35$, in the present case.

2.7.6 A semi-quantitative explication of the possible behaviours

On general grounds, the rounding of the transition due to disorder in the present model should be mainly attributable to the presence of *rare regions* in the sequences. This is expected in particular from the theoretical results [56], although the key role of the *atypical events* in the proof of disorder relevance in [20, 21, 6] decisively appears to point towards the same direction, in different terms. Accordingly, we attempt to roughly quantify the *rare region* contribution in the finite size behaviour of the system, to better understand the way in which relevance of disorder becomes manifest, starting from the finite size level.

Within this context, for the sake of completeness, we start the discussion here by recalling in some detail a hypothesis originally made in [37], in which a very simplified phenomenological picture was proposed to this aim. In fact, it can be noticed that, in the PS model, both the temperature and the disorder enter only in the exponential terms in the partition function. This last one can be written in a form similar to (2.1), with the difference that we are now referring to the considered DSAWPS model. Hence, with μ the on-lattice connectivity constant and c_p the *renormalized* loop exponent of SAWs on a $3D$ (cubic) lattice (*i.e.*, $c_p = 2.15$), n_h the segment length, l_h the loop length and $2r = \mathcal{N} - 2\sum_{h=1}^{k}(n_h + l_h)$, one has:

$$Z_{\mathcal{N}}^{DSAWPS} = \mu^{2\mathcal{N}} e^{(\beta\epsilon_1 - \log\mu)} \sum_{\{n,l\}} \prod_{h=0}^{k} \left[\prod_{\{\sigma_i\}_h} e^{\sigma_i(\beta\epsilon_i - \log\mu)} \right] \left(\frac{1}{2l_h} \right)^{c_p} . \tag{2.25}$$

Here, once again, the sum runs over all the partitions in k segments of total length n and k loops of total length $2l$, with $2(n+l) \leq \mathcal{N}$. In this formula, we have moreover introduced the variables $\{\sigma_i\}_h$, with $\sigma_i \in \{0,1\}$ and $\sigma_i = 1$, in a given configuration, *if and only if* the bases in position i belong to the h-th segment of the considered partition (that has length n_h). Such a writing allows us to highlight just the exponential terms, $\pi_i(h) = e^{\sigma_i(\beta\epsilon_i - \log\mu)}$, with $\sigma_i \in \{\sigma_i\}_h$ (by taking implicitly $\sigma_1 = 1$).

Moreover, we recall that, in the pure system, the transition occurs around the temperature $\mathcal{T}_{c,p} \sim \epsilon/\log\mu$, at which the energetic contribution for the two bound chains is of the same order as the entropic loss. In the presence of disorder, for a given sequence, one knows that the multi-step behaviour of the order parameter that is observed in experimental DNA denaturation curves mainly results from the presence of regions with different local contents in terms of GC to AT ratios. Accordingly, these regions have different local melting temperatures.

In the simplest extreme case, one can imagine two regions A and B, of about the same length L, completely dominated by AT and GC compositions, respectively. In such a situation, the local transition in region A is driven by ϵ_{AT} energies, with local critical temperature $\mathcal{T}_{c,loc}(A) \sim \mathcal{T}_{c,AT} \sim \epsilon_{AT}/\log\mu$, whereas the local transition in region B is associated with the higher local critical temperature $\mathcal{T}_{c,loc}(B) \sim \mathcal{T}_{c,GC} \sim \epsilon_{GC}/\log\mu$. Therefore, $\mathcal{T}_{c,loc}(B) \sim R\mathcal{T}_{c,loc}(A)$, with $R = \epsilon_{GC}/\epsilon_{AT}$.

In this illustrative example, for a given temperature, the contributions in a configuration (*i.e.*, of a given succession $\in \{n,l\}$) in the partition function to the corresponding total exponential factor:

$$\begin{aligned} \pi^{tot} &= \prod_h \prod_{\{\sigma_i\}_h} \pi_i(h) = \left[\prod_h \prod_{i\in A} \pi_i(h) \right] \left[\prod_h \prod_{i\in B} \pi_i(h) \right] \left[\prod_h \prod_{i\notin A\cup B} \pi_i(h) \right] = \\ &= \pi_A^{tot}\pi_B^{tot} \left[\prod_h \prod_{i\notin A\cup B} \pi_i(h) \right] \end{aligned} \tag{2.26}$$

will be significantly different for A and B regions. The configuration in which this effect is the most evident is clearly the one in which both the thermal variables in the A region and the ones in the B region belong to segments (*i.e.*, $\sigma_i = 1\ \forall i \in A \cup B$). In such a simple case, one obtains, for $\mathcal{T} = \mathcal{T}_{c,loc}(A)$:

$$\pi_A^{tot} = \prod_{i\in A} e^{(\beta_{c,loc}(A)\epsilon_{AT} - \log\mu)} = e^{[L(\beta_{c,loc}(A)\epsilon_{AT} - \log\mu)]} \sim O(1) \tag{2.27}$$

$$\pi_B^{tot} = \prod_{i\in B} e^{(\beta_{c,loc}(A)\epsilon_{GC} - \log\mu)} = e^{[L(\beta_{c,loc}(A)\epsilon_{GC} - \log\mu)]} \sim e^{L/x^{min}}, \tag{2.28}$$

In fact, whereas $(\beta_{c,loc}(A)\epsilon_{AT} - \log\mu) \sim 0$, one has:

$$(\beta_{c,loc}(A)\epsilon_{GC} - \log\mu) \sim (R-1)\log\mu \equiv \frac{1}{x^{min}}, \tag{2.29}$$

that defines the parameter x^{min}. Analogously, by moving forward the reasoning in order to make it more rigorous, one can imagine the same situation at the different temperature $\mathcal{T} = \mathcal{T}_{c,loc}(B)$:

$$\pi_A^{tot} = \prod_{i\in A} e^{(\beta_{c,loc}(B)\epsilon_{AT} - \log\mu)} = e^{[L(\beta_{c,loc}(B)\epsilon_{AT} - \log\mu)]} \sim e^{-L/x^{max}} \tag{2.30}$$

$$\pi_B^{tot} = \prod_{i\in B} e^{(\beta_{c,loc}(B)\epsilon_{GC} - \log\mu)} = e^{[L(\beta_{c,loc}(B)\epsilon_{GC} - \log\mu)]} \sim O(1), \tag{2.31}$$

In this opposite case, whereas $(\beta_{c,loc}(B)\epsilon_{GC} - \log\mu) \sim 0$, one has:

$$(\beta_{c,loc}(B)\epsilon_{AT} - \log\mu) \sim -\frac{R-1}{R}\log\mu \equiv -\frac{1}{x^{max}}, \tag{2.32}$$

that defines the parameter x^{max}.

From these expressions it is possible to argue that, the larger the value of L with respect to $x \in [x^{min}, x^{max}]$, the more the effect of disorder will be *felt* by the finite size system, *i.e.*, the difference between the weights of configurations corresponding to closed A and B regions in the partition function will be higher. On the other hand, the probability for such an extreme case in a particular sequence of length $\mathcal{N}$ is quite small, as we are going to approximately quantify. One can expect that, in particular in the considered case of a binomial distribution (see (2.4)), this probability, though approaching 1 in the thermodynamic limit for any finite L, becomes rapidly negligible with increasing L for fixed chain length $\mathcal{N}$. Following these considerations, the $\mathcal{N}$-value necessary for the finite size system to feel the effect of disorder, hence for observing rare regions with $L \gg x$, could be not reachable for large x values.

To conclude the present phenomenological picture by attempting to understand the role of the loop entropic effect, one can impose that, at the temperature $\mathcal{T}_{c,loc}(B)$, the weight of the configuration associated with region A in the closed state be significantly smaller than that of the configuration associated with an open region corresponding to a single loop (of size L), getting the condition $L/x \gg c_p \log L$. Correspondingly, one can argue that, with increasingly larger c_p values, increasingly larger rare region lengths will be necessary for observing cooperative melting behaviour at different temperatures. Moreover, the considered extreme case seems particularly appropriate for a qualitative description of the case $c_p > 2$. Here, the first order character of the transition in the pure system and the corresponding favored formation of small loops suggests that larger differences of local AT to GC content ratio are necessary for obtaining different local melting temperatures. Thus, to evaluate the minimal length L of "all AT" or "all GC" regions that is necessary for the effect of disorder being evident is even more a reasonable starting point.

In the present work, we apply this phenomenological picture in order to obtain an evaluation of such a minimal "rare region length" L, on the basis of the present numerical evaluation of $\mathcal{N}^*$.

Summarizing, qualitatively speaking, from the finite size behaviour point of view, the considered model appears to be characterized in general by a slow approach to the asymptotic regime, and the evidence for the effect of disorder (leading to the predicted smooth transition described by a single correlation length) appears to be related to the quite sudden appearance, at $\mathcal{N} \sim \mathcal{N}^*$, of multi-step behaviour in the order parameter in a significant fraction of the sequences. This feature supports our phenomenological picture, hence the hypothesis that the order of magnitude of the chain length at which the effect of disorder becomes evident is just to be related to the presence in such sequences of *large enough rare regions*, that would then be already occurring, at $\mathcal{N} \sim \mathcal{N}^*$, with non negligible probability.

From such a perspective, the parameter x that we introduced is to be interpreted as a measure of the *effective* disorder strength, linking the concept of large enough rare regions to the values of $R = \epsilon_{GC}/\epsilon_{AT}$ and $\log \mu$ in the model.

Accordingly, we are led to the combinatorial problem of evaluating the probability, $P(L, \mathcal{N})$, to observe, in a sequence of total length $\mathcal{N}$, a subsequence of consecutive base pairs of the same kind of length at least L. Such an evaluation could appear to be rather simple. However even the obtention of the exact solution for the binomial quenched disordered variables distribution given by (2.4) appears to be far from trivial. Indeed, this is the subject of our subsequent investigation, whose preliminary results we will discuss in the following Section 2.8. For the sake of consistency, we resort here to the approximation used in [32]:

$$P(L, \mathcal{N}) \sim (\mathcal{N} - L + 1)\, 2^{-L}. \tag{2.33}$$

Such an approximation is exact for $L = \mathcal{N}$, but clearly invalid for small L values, leading to probabilities higher than 1. Nonetheless, it can be considered that this approximation provides a reasonable basis for the present analysis, as we are mainly interested to capture the order of magnitude of the quantity (we checked numerically, by exact computations, that this is roughly the case, in a large L range, already for small $\mathcal{N}$ values).

As outlined above, at the crossover it is expected to find large enough rare regions with a non-negligible probability. Here we set this probability to the value 0.5, based on the observation that the order parameter displays multi-step behaviour at $\mathcal{N} \sim \mathcal{N}^*$ in a fraction $\sim 0.3 - 0.7$ of the studied sequences. Accordingly, by applying (2.33) with $P(\mathcal{N}^*, L^*) = 0.5$ (hence finding numerically the solution $L^*(\mathcal{N}^*)$ of $(\mathcal{N}^* - L^* + 1)2^{-L^*} = 0.5$), we get an estimation of the crossover rare region length $L^* \sim 11 - 14.5$, in correspondence with the crossover chain length $\mathcal{N}^* \sim 10^3 - 10^4$ that characterizes the present model.

Then, by assuming that the phenomenological picture does indeed capture the basic physics of the problem, and by noticing that one has $x = x_{CY} \in [0.65, 1.3]$ in the present case, it is in general easy to predict the behaviours of $L^*(x)$ and of the corresponding crossover chain length $\mathcal{N}^*(x)$ as functions of the parameter x in models *à la* PS for DNA denaturation transition with $c_p = 2.15$. Indeed, within the proposed qualitative picture, in order for the effect of disorder to be equally manifest in different models, it is expected to be necessary for the underlying ratio values $L^*(x)/x$ (of the crossover rare region length to the parameter x) to be similar. Such a condition can be expressed as $L^*(x) = L^*(x_{CY})x/x_{CY}$. It is then possible to obtain $\mathcal{N}^*(x)$ by solving once again (2.33) with $P(\mathcal{N}^*, L^*) = 0.5$.

In this way, for the PS model with $c_p = 2.15$ in [38, 39], we find extremely large $\mathcal{N}^*$ values of order larger than 10^{30}. Indeed, the link energies and the connectivity constant used in [38, 39] lead to $x = x_{GM} \in [15, 16.4]$. Taking also into account the slightly different law for the allowed coupling adopted in that study, the effective value of the parameter x_{GM} is in fact expected to be still larger. Even though relying on rough evaluations, it appears therefore reasonable to consider that with the conditions in [38, 39] it would be impossible, in practice, to reach the asymptotic regime. In such a context, the PS model with $c_p = 2.15$ in [38, 39] is expected to behave in agreement with a pseudo first order transition, ruled by two correlation lengths (corresponding to typical and average

quantities; namely with $\nu_{r,1} = \nu_p = 1$ and $\nu_{r,2} = 2$), independently from the considered observable and even well beyond the (already extremely large) reached chain lengths (up to $\mathcal{N} = 2 \cdot 10^6$).

Thus, in the context of the analysis above, it appears possible to reconcile the two different pictures emerging from the previous numerical studies [36, 37, 38, 39]. Further, in such a context, the scenario associated with the pseudo first order transition appears meaningful to describe the finite-$\mathcal{N}$ behaviour in the presence of *weak* disorder.

Finally, it is worth noting that the x_{CY} of the present model, that pertains to the region associated with large finite size effects (yet possible to study), is definitely closer to the still *smaller* $x_{exp} \in [0.18, 0.2]$ associated with the values of the coupling energies and of $\log \mu$ typically adopted for comparisons with experimental melting curves. In detail, the value of R adopted for comparison with experiments is essentially the same ($R \simeq 1.1$) as the one in [38, 39], with, however, underlying link energies more than one order of magnitude smaller and a $\log \mu$ value more than one order of magnitude larger. The simple approximation of $P(L, \mathcal{N})$ that we considered here may not be sufficient for precise predictions of L^*_{exp} and $\mathcal{N}^*_{exp}$ values. However, it seems reasonable to expect crossover rare region length of the order of a few base pairs, and a corresponding relatively small crossover chain length. From this point of view, the present results on the studied PS model with $c_p = 2.15$ suggest that both self-avoidance and disorder could play a key role in experimental DNA behaviour.

Apart from the possible effect of neglecting correlations in the sequences, which are well known to be present in DNA molecules, a throughout understanding of this phenomenon would also require to better clarify the role of an additional parameter that, as advanced, is usually introduced in PS model: the cooperativity factor, σ_0, that we set equal to 1 to reproduce the behaviour of the on-lattice DSAWDNA, but that should instead make more unlikely both the opening and the closing of loops, hence enhance just the cooperativity. Indeed, in order to reproduce correctly the experimental curves, usually very small values, $\sigma_0 \sim 10^{-4} - 10^{-5}$, are adopted for this parameter. It would be meaningful to further characterize the influence of this parameter, that should not influence the behaviour at the critical point, on the finite size behaviour of the system (a value of $\sigma_0 \neq 1$ can be expected to roughly correspond to the introduction of an additional, not diverging, correlation length $\xi_{\sigma_0} \sim 1/\sigma_0$).

Anyway, from this last point of view, it was observed in [47] that, when taking $c_p \simeq 2.15$, one could use larger (of about a factor 10) σ_0 value. This appears as a further confirmation that to correctly take into account the importance of rare regions allows to give a less important role to such a cooperativity parameter whose introduction, in the PS model, is basically justified from the necessity to reproduce the effects of the double helix structure of DNA molecules.

2.8 On the probability of obtaining at least L heads by tossing $\mathcal{N}$ times a fair coin

We are interested in exactly calculating the probability of observing a rare region of length at least L in a sequence of length $\mathcal{N}$, for a uniform binomial distribution of the two possible linking energies. By definition, such a probability, $P(L, \mathcal{N})$, is the same as the one to obtain at least L consecutive heads by tossing $\mathcal{N}$ times a fair coin.

Indeed, by definition of a fair coin, the outcome of a single toss can be head ($\bullet$) or tail ($\circ$) with equal probability 1/2. In other words, the outcome of each toss is an independent identically distributed random variable that follows the uniform binomial distribution. Therefore, the outcome of $\mathcal{N}$ tosses is a given sequence of $\bullet$ and $\circ$, $s(\mathcal{N})$, *i.e.*, by labeling $\circledast$ the variable that can take each of the two possible values:

$$s(\mathcal{N}) \in \mathcal{S}(\mathcal{N}) \equiv \{\overbrace{\circledast \cdots \circledast}^{\mathcal{N}}\}. \tag{2.34}$$

The cardinality of the set $\mathcal{S}(\mathcal{N})$, $|\mathcal{S}(\mathcal{N})|$, hence the total number of possible sequences, is equal to $2^{\mathcal{N}}$ and all the possible different sequences have the same probability $1/2^{\mathcal{N}}$.

Let us introduce the subset $\mathcal{S}_L(\mathcal{N}) \subset \mathcal{S}(\mathcal{N})$ of the sequences $s(\mathcal{N}) \in \mathcal{S}(\mathcal{N})$ with *at least* L consecutive $\bullet$. By definition, one has:

$$P(L,\mathcal{N}) = \frac{|\mathcal{S}_L(\mathcal{N})|}{|\mathcal{S}(\mathcal{N})|}. \tag{2.35}$$

2.8.1 Calculation of $P(L,\mathcal{N})$ from the L-step Fibonacci numbers

Let us define the complementary subset, $\overline{\mathcal{S}_L(\mathcal{N})} \equiv \mathcal{S}(\mathcal{N}) \setminus \mathcal{S}_L(\mathcal{N})$, *i.e.*, the set of the sequences that contain less than L consecutive $\bullet$, whose cardinality is $|\overline{\mathcal{S}_L(\mathcal{N})}| \equiv 2^{\mathcal{N}} - |\mathcal{S}(\mathcal{N})|$. Equivalently:

$$|\mathcal{S}_L(\mathcal{N})| = 2^{\mathcal{N}} - |\overline{\mathcal{S}_L(\mathcal{N})}| \tag{2.36}$$

Clearly, $\forall \mathcal{N} \in \mathbb{N}$, there is a single sequence that does not contain any $\bullet$, hence $|\overline{\mathcal{S}_1(\mathcal{N})}| = 1$ and

$$|\mathcal{S}_1(\mathcal{N})| = 2^{\mathcal{N}} - 1. \tag{2.37}$$

In order to calculate $|\mathcal{S}_2(\mathcal{N})|$, let us start by writing explicitly the sets $\overline{\mathcal{S}_2(\mathcal{N})}$ for small $\mathcal{N} \leq 4$ values:

$$\begin{aligned}
\overline{\mathcal{S}_2(1)} &= \{\circ, \bullet\} \\
\overline{\mathcal{S}_2(2)} &= \{\circ\,\circ, \circ\,\bullet, \bullet\,\circ\} \\
\overline{\mathcal{S}_2(3)} &= \{\circ\,\circ\,\circ, \circ\,\circ\,\bullet, \circ\,\bullet\,\circ, \bullet\,\circ\,\circ, \bullet\,\circ\,\bullet\} \\
\overline{\mathcal{S}_2(4)} &= \{\circ\,\circ\,\circ\,\circ, \circ\,\circ\,\circ\,\bullet, \circ\,\circ\,\bullet\,\circ, \circ\,\bullet\,\circ\,\circ, \circ\,\bullet\,\circ\,\bullet, \bullet\,\circ\,\circ\,\circ, \bullet\,\circ\,\circ\,\bullet, \bullet\,\circ\,\bullet\,\circ\}.
\end{aligned} \tag{2.38}$$

First of all, we notice that (2.36) cannot be satisfied for $\mathcal{N} = 1$, which is obvious, since $L > \mathcal{N}$. It is instead satisfied for $\mathcal{N} \geq 2$ (for instance $|\mathcal{S}_2(2)| = |\{\bullet\bullet\}| = 1 = 2^2 - 3$).

Then, one finds that:

$$\begin{aligned}
\overline{\mathcal{S}_2(3)} &= [(\circ)\{\circ\,\circ, \circ\,\bullet, \bullet\,\circ\}] \bigcup [(\bullet\,\circ)\{\circ, \bullet\}] = \left[(\circ)\overline{\mathcal{S}_2(2)}\right] \bigcup \left[(\bullet\,\circ)\overline{\mathcal{S}_2(1)}\right] \\
\overline{\mathcal{S}_2(4)} &= [(\circ)\{\circ\,\circ\,\circ, \circ\,\circ\,\bullet, \circ\,\bullet\,\circ, \bullet\,\circ\,\circ, \bullet\,\circ\,\bullet\}] \bigcup [(\bullet\,\circ)\{\circ\,\circ, \circ\,\bullet, \bullet\,\circ\}] = \\
&= \left[(\circ)\overline{\mathcal{S}_2(3)}\right] \bigcup \left[(\bullet\,\circ)\overline{\mathcal{S}_2(2)}\right].
\end{aligned} \tag{2.39}$$

The result:

$$\overline{\mathcal{S}_2(\mathcal{N}+1)} = \left[(\circ)\overline{\mathcal{S}_2(\mathcal{N})}\right] \bigcup \left[(\bullet\,\circ)\overline{\mathcal{S}_2(\mathcal{N}-1)}\right], \tag{2.40}$$

is true $\forall \mathcal{N} \geq 2$, since, in this case: i) $\forall s(\mathcal{N}) \in \overline{\mathcal{S}_2(\mathcal{N})}$ one can obtain $s(\mathcal{N}+1) \in \overline{\mathcal{S}_2(\mathcal{N}+1)}$ (hence, by avoiding the subsequence $\bullet\,\bullet$) by adding a $\circ$ at its beginning (or, respectively, at its end); ii) $\forall s(\mathcal{N}-1) \in \overline{\mathcal{S}_2(\mathcal{N}-1)}$ one can obtain $s(\mathcal{N}+1) \in \overline{\mathcal{S}_2(\mathcal{N}+1)}$ (hence, by avoiding the subsequence $\bullet\,\bullet$) by adding a $\bullet\,\circ$ at its beginning (or, respectively, a $\circ\,\bullet$ at its end); iii) any sequence $s(\mathcal{N}+1) \in \overline{\mathcal{S}_2(\mathcal{N}+1)}$ can be obtained by one of the previous two ways and the sequences obtained with such a method are all different among themselves.

Therefore, we obtain:

$$|\overline{\mathcal{S}_2(\mathcal{N}+1)}| = |\overline{\mathcal{S}_2(\mathcal{N})}| + |\overline{\mathcal{S}_2(\mathcal{N}-1)}| \qquad \text{with } |\overline{\mathcal{S}_2(1)}| = 2;\ |\overline{\mathcal{S}_2(2)}| = 3, \tag{2.41}$$

implying that the sequence of the $|\overline{\mathcal{S}_2(\mathcal{N})}|$ is the sequence of the Fibonacci numbers [70], $F_2(n)$, with $n = \mathcal{N} + 2$.

Let us continue by writing explicitly also the sets $\overline{\mathcal{S}_3(\mathcal{N})}$ for small $\mathcal{N} \leq 4$ values:

$$\overline{\mathcal{S}_3(1)} = \{\circ, \bullet\}$$

$$\begin{aligned}
\overline{\mathcal{S}_3(2)} &= \{\circ\circ, \circ\bullet, \bullet\circ, \bullet\bullet\} \\
\overline{\mathcal{S}_3(3)} &= \{\circ\circ\circ, \circ\circ\bullet, \circ\bullet\circ, \circ\bullet\bullet, \bullet\circ\circ, \bullet\circ\bullet, \bullet\bullet\circ\} \\
\overline{\mathcal{S}_3(4)} &= \{\circ\circ\circ\circ, \circ\circ\circ\bullet, \circ\circ\bullet\circ, \circ\circ\bullet\bullet, \circ\bullet\circ\circ, \circ\bullet\circ\bullet, \\
&\quad \circ\bullet\bullet\circ, \bullet\circ\circ\circ, \bullet\circ\circ\bullet, \bullet\circ\bullet\circ, \bullet\circ\bullet\bullet, \bullet\bullet\circ\circ, \bullet\bullet\circ\bullet\}.
\end{aligned} \tag{2.42}$$

It appears evident that the recurrence relation (2.41) can be generalized to higher L values. In detail, by following the same way of reasoning as previously, we can prove that, $\forall\mathcal{N} \geq 3$:

$$\begin{aligned}
|\overline{\mathcal{S}_3(\mathcal{N}+1)}| &= |\overline{\mathcal{S}_3(\mathcal{N})}| + |\overline{\mathcal{S}_3(\mathcal{N}-1)}| + |\overline{\mathcal{S}_3(\mathcal{N}-2)}|, \\
&\quad \text{with } |\overline{\mathcal{S}_3(1)}| = 2;\ |\overline{\mathcal{S}_3(2)}| = 4;\ |\overline{\mathcal{S}_3(3)}| = 7,
\end{aligned} \tag{2.43}$$

since:

$$\overline{\mathcal{S}_3(\mathcal{N}+1)} = (\circ)\overline{\mathcal{S}_3(\mathcal{N})} + (\bullet\,\circ)\overline{\mathcal{S}_3(\mathcal{N}-1)}| + (\bullet\ \bullet\ \circ)\overline{\mathcal{S}_3(\mathcal{N}-2)}|. \tag{2.44}$$

Correspondingly, the sequence of the $|\overline{\mathcal{S}_3(\mathcal{N})}|$ is the sequence of the 3-step Fibonacci numbers, $F_3(n)$, with $n = \mathcal{N}+2$.

In order to further generalize, we notice that, $\forall\mathcal{N}$, there is a single sequence that contain only $\bullet$, too. Therefore, by using (2.36), one has:

$$|\overline{\mathcal{S}_{\mathcal{N}}(\mathcal{N})}| = 2^{\mathcal{N}} - 1. \tag{2.45}$$

In detail, we get $\forall\mathcal{N} \geq 4$:

$$\begin{aligned}
|\overline{\mathcal{S}_4(\mathcal{N}+1)}| &= |\overline{\mathcal{S}_4(\mathcal{N})}| + |\overline{\mathcal{S}_4(\mathcal{N}-1)}| + |\overline{\mathcal{S}_4(\mathcal{N}-2)}| + |\overline{\mathcal{S}_4(\mathcal{N}-3)}|, \\
&\quad \text{with } |\overline{\mathcal{S}_4(1)}| = 2;\ |\overline{\mathcal{S}_4(2)}| = 4;\ |\overline{\mathcal{S}_4(3)}| = 8;\ |\overline{\mathcal{S}_4(4)}| = 15;
\end{aligned} \tag{2.46}$$

and, by generalizing, $\forall\mathcal{N} \geq L$, $\forall L \geq 2$:

$$|\overline{\mathcal{S}_L(\mathcal{N}+1)}| = \sum_{k=0}^{L-1} |\overline{\mathcal{S}_L(\mathcal{N}-k)}| \text{ with } |\overline{\mathcal{S}_L(k)}| = 2^k \ \ \forall 1 \leq k < L;\ |\overline{\mathcal{S}_L(L)}| = 2^L - 1. \tag{2.47}$$

In other words, the sequence $|\overline{\mathcal{S}_L(\mathcal{N})}|$ is the sequence of the L-step Fibonacci numbers [71, 72, 73], $F_L(n)$, with $n = \mathcal{N}+2$. Notice that one can (arbitrarily but appropriately) define also the first terms in (2.47), *i.e.*, $|\overline{\mathcal{S}_L(0)}| = |\overline{\mathcal{S}_L(-1)}| = 1$ and $|\overline{\mathcal{S}_L(-k)}| = 0$ for $k \in [2, (L-2)]$, in such a way that, in particular, $F_L(0) = 0, F_L(1) = F_L(2) = 1\ \forall L$.

In conclusion, for $L \leq (\mathcal{N}-1)$, by virtue of (2.36) and (2.47), we obtain $|\mathcal{S}_L(\mathcal{N})|$, hence $P(L,\mathcal{N})$.

We outline that such a relation between the $P(L,\mathcal{N})$ and the L-step Fibonacci numbers appears quite frequently in the recent literature. For instance, the sequences $|\mathcal{S}_2(\mathcal{N})|$, $|\mathcal{S}_3(\mathcal{N})|$, $|\mathcal{S}_4(\mathcal{N})|$ and $|\mathcal{S}_5(\mathcal{N})|$ correspond to the sequences *A008466*, *A050231*, *A050232* and *A050233* in [74], respectively. In detail, in [74], these sequences are defined just as the number of $\mathcal{N}$-tosses having a run of L or more heads for a fair coin (with L=2, 3, 4 and 5, respectively) and they are related to the corresponding sequences of the L-step Fibonacci numbers. Nonetheless, we find that the present sketch of the derivation of the result makes the work more self-consistent.

Moreover, we also notice that the L-step Fibonacci numbers are examples of homogeneous linear recurrences of k-th degree with constant coefficients, whose characteristic polynomials $p_L(x)$ are known [70]. On the other hand, to deepen such a subject is outside the aims of the present work, since it would not contribute to the numerical implementation of the calculation of $P(L,\mathcal{N})$.

2.8.2 Calculation of $P(L,\mathcal{N})$ for $L \geq \lceil \mathcal{N}/2 \rceil$

Let us furthermore introduce the subset of sequences in which the longest subsequence made by consecutive $\bullet$ contains *exactly* L of them, *i.e.*, the subset of sequences with *at most exactly* L consecutive $\bullet$, $\mathcal{S}^*_L(\mathcal{N})$.

One has, $\forall L \in [1,\, \mathcal{N}]$, $\forall \mathcal{N} \geq 1$, :

$$\mathcal{S}_L(\mathcal{N}) = \bigcup_{k=L}^{\mathcal{N}} \mathcal{S}^*_k(\mathcal{N}), \tag{2.48}$$

as can be verified by induction. It is in particular, clearly:

$$|\mathcal{S}_{\mathcal{N}}(\mathcal{N})| = |\mathcal{S}^*_{\mathcal{N}}(\mathcal{N})| = 1. \tag{2.49}$$

For $\mathcal{N} - 2 \geq L \geq \lceil \mathcal{N}/2 \rceil$, we find furthermore the relation:

$$\begin{aligned} \mathcal{S}^*_L(\mathcal{N}) \;=\; & \left(\bigcup_{h=0}^{\mathcal{N}-(L+2)} \{ \overbrace{\circledast \cdots \circledast}^{h} \;\; \overbrace{\circ \bullet \cdots \bullet \circ}^{L+2} \;\; \overbrace{\circledast \cdots \circledast}^{\mathcal{N}-(L+2)-h} \} \right) \bigcup \\ & \bigcup \left(\{ \overbrace{\bullet \cdots \bullet \circ}^{L+1} \; \overbrace{\circledast \cdots \circledast}^{\mathcal{N}-(L+1)} \} \right) \bigcup \left(\{ \overbrace{\circledast \cdots \circledast}^{\mathcal{N}-(L+1)} \; \overbrace{\circ \bullet \cdots \bullet}^{L+1} \} \right). \end{aligned} \tag{2.50}$$

In fact, in the case $L \geq \lceil \mathcal{N}/2 \rceil$, there is no possibility for a subsequence of length larger than L in the remaining part of the sequence.

In order to better understand the relation (2.50), we continue by calculating once again explicitly $\mathcal{S}^*_L(\mathcal{N})$ for small $\mathcal{N} \leq 4$.

$$\begin{aligned} \mathcal{S}^*_1(2) &= \{\circ\bullet, \bullet\circ\} \\ \mathcal{S}^*_2(3) &= \{\bullet\bullet\circ, \circ\bullet\bullet\} \\ \mathcal{S}^*_3(4) &= \{\bullet\bullet\bullet\circ, \circ\bullet\bullet\bullet\} \end{aligned} \tag{2.51}$$

Indeed, $\forall \mathcal{N} \geq 2$, $|\mathcal{S}^*_{\mathcal{N}-1}(\mathcal{N})| = 2$. This can be understood by observing that, in order to get a sequence with $\mathcal{N}-1$ consecutive $\bullet$, one can only change from $\bullet$ to $\circ$ a variable at one of the two extremities of the single sequence made by $\mathcal{N}$ consecutive $\bullet$.

One can also obtain that, $\forall \mathcal{N} \geq 4$, $|\mathcal{S}^*_{\mathcal{N}-2}(\mathcal{N})| = 5$. In detail, for $\mathcal{N} = 4$, one has $|\mathcal{S}^*_2(4)| = 5$, since:

$$\mathcal{S}^*_2(4) = \{\bullet\bullet\circ\circ, \bullet\bullet\circ\bullet, \circ\bullet\bullet\circ, \bullet\circ\bullet\bullet, \circ\circ\bullet\bullet\} \tag{2.52}$$

For $\mathcal{N} > 4$, such a result is given from the two possible sequences $\bullet\cdots\bullet\circ\circledast$ plus the two possible sequences $\circledast\circ\bullet\cdots\bullet$ plus the single sequence $\circ\bullet\cdots\bullet\circ$.

On the basis of the same kind of reasoning, one has:

$$\begin{aligned} &\forall \mathcal{N} \geq 6,\;\; |\mathcal{S}^*_{\mathcal{N}-3}(\mathcal{N})| = 12 \qquad (\{\bullet\cdots\bullet\circ\circledast\circledast\} \cup \{\circ\bullet\cdots\bullet\circ\circledast\} \cup \{\circledast\circ\bullet\cdots\bullet\circ\} \cup \{\circledast\circledast\circ\bullet\cdots\bullet\}) \\ &\forall \mathcal{N} \geq 8,\;\; |\mathcal{S}^*_{\mathcal{N}-4}(\mathcal{N})| = 28 \qquad (\{\bullet\cdots\bullet\circ\circledast\circledast\circledast\} \cup \{\circ\bullet\cdots\bullet\circ\circledast\circledast\} \cup \{\circledast\circ\bullet\cdots\bullet\circ\circledast\} \cup \\ &\qquad\qquad \cup\{\circledast\circledast\circ\bullet\cdots\bullet\circ\} \cup \{\circledast\circledast\circledast\circ\bullet\cdots\bullet\}) \end{aligned} \tag{2.53}$$

By generalizing, we obtain the following result, that is indeed in agreement with the relation between subsets in (2.50):

$$\forall \mathcal{N} \geq 2k,\;\; |\mathcal{S}^*_{\mathcal{N}-k}(\mathcal{N})| = 2^{k-2}(k+3). \tag{2.54}$$

We have already seen that this result is true for $0 < k \leq 4$. First of all, we underline that the number of sequences counted by $|\mathcal{S}^*_{\mathcal{N}-k}(\mathcal{N})|$ in the case $\mathcal{N} = \mathcal{N}_0 = 2k$ is the same as the number

of sequences counted by $|\mathcal{S}^*_{\mathcal{N}-k}(\mathcal{N})|$ for $\mathcal{N} > 2k$, since the last are the ones and the only ones in which the subsequences $\bullet\cdots\bullet$ of length $\mathcal{N}_0/2$ are replaced by subsequences of consecutive $\bullet$ of length $\mathcal{N} - k > \mathcal{N}_0/2$. In order to prove the result $\forall \mathcal{N} = 2k$, let us note that: i) the contributing sequences that start or end with the subsequence $\bullet\cdots\bullet$ are $2 \cdot 2^{\mathcal{N}-k-1} = 4 \cdot 2^{k-2}$, since both of them contain $\mathcal{N} - k - 1$ free variables $\circledast$; ii) all the other sequences contain one more fixed variable $\circ$, hence their number is proportional to $2^{\mathcal{N}-k-2} = 2^{k-2}$. In detail, their number is exactly $[\mathcal{N}-(\mathcal{N}-k)-1]\cdot 2^{k-2} = (k-1)\cdot 2^{k-2}$, since one can choose in $[\mathcal{N} - (\mathcal{N} - k) - 1]$ ways the origin of the $\circ\bullet\cdots\bullet\circ$ subsequence in each case. Correspondingly, one has $S^*_k(\mathcal{N}) = 2^{k-2}(k+3)$. Clearly, such a derivation can be done only for $k > 2$, but we know that the result is correct in particular for $k = 1$ and $k = 2$.

Summarizing, from (2.54) and (2.48), by moreover recalling that $|\mathcal{S}^*_{\mathcal{N}}(\mathcal{N})| = 1$, given a sequence of length $\mathcal{N}$ we obtain the following result on $|S_L(\mathcal{N})|$, that is correct for $L \geq \lceil \mathcal{N}/2 \rceil$:

$$|\mathcal{S}_L(\mathcal{N})| = 1 + \sum_{h=1}^{k} 2^{h-2}(h+3) = 2^{k-1}(k+2) \text{ with } k = \mathcal{N} - L. \tag{2.55}$$

Importantly, both the sequence $|\mathcal{S}^*_k|$ ($\{1, 2, 5, 12, 28, 64, \dots\}$) and the sequence $|\mathcal{S}_k|$ ($\{1, 3, 8, 20, 48, 112, \dots\}$) are well described in the literature (they correspond to the sequences *A045623* and *A001792* in [74], respectively). Nonetheless, their key role in the calculation of the present $P(L,\mathcal{N})$, in particular from the point of view of implementing the numerical calculation for *finite but large* $\mathcal{N}$ value, was not yet outlined, to our knowledge.

On the other hand, we know that (2.36)-(2.47), *i.e.*, the relation between $P(L,\mathcal{N})$ and the L-step Fibonacci numbers, are correct $\forall\ 2 \leq L \leq \mathcal{N}$. Therefore, for $L \geq \lceil \mathcal{N}/2 \rceil$, we obtain the following result:

$$2^{\mathcal{N}} - |\overline{\mathcal{S}_L(\mathcal{N})}| = 2^{\mathcal{N}-L-1}(\mathcal{N} - L + 2), \tag{2.56}$$

that is equivalent to:

$$F_L(n) = 2^{n-2} - 2^{n-L-3}(n-L) \text{ for } L \leq n - 2 \leq 2L. \tag{2.57}$$

In fact, this result holds for the L-step Fibonacci numbers, as stated in [75].

2.8.3 Preliminary results on the evaluation of $\mathcal{N}^*$

On the basis of the exact knowledge of $P(L,\mathcal{N})$, one can in principle both refine the given evaluations of $\mathcal{N}^*$ and obtain the behaviour of this quantity as a function of the disordered strength x, hence of the precise parameter values, in the PS model with $c_p = 2.15$.

Here we limit ourselves to present in [Fig. 2.16] the refined numerical evaluation of the minimal rare region length, L^*, for the considered PS model, that is $L^* \simeq 10 - 11$. As in the previous Subsection 2.7.6, L^* is defined as the abscissa that corresponds to $P(L^*,\mathcal{N}^*) = 0.5$, whereas, more precisely than previously, we use 10^3 and $5 \cdot 10^3$ as our best estimations (from below and from above, respectively) of the crossover chain length $\mathcal{N}^*$ in the considered PS model.

We moreover notice that, generally, $L^* \ll \mathcal{N}^*/2$, since $P(L,\mathcal{N})$ appears characterized by an initial region in which one finds a nearly constant $P(L,\mathcal{N}) \simeq 1$, that is followed by a rapid decrease towards zero at still relatively small $L \ll \mathcal{N}/2$ values. In fact, we observed this kind of behaviour already for relatively small $\mathcal{N}$ values, and we observed that the length of the initial region increases very slowly with respect to $\mathcal{N}$, too.

Correspondingly, the interesting part of the $P(L,\mathcal{N})$ curve is the one that can be in principle calculated by means of the Fibonacci L-step numbers. Indeed, this is expected to allow a meaningful evaluation of $\mathcal{N}^*$ in the case of the parameter values that are used when comparing with experimental data.

On the other hand, our preliminary results seem to show that the difference between the correct $P(L,\mathcal{N})$ and the closed formula (2.55), that holds for $L \geq \lceil \mathcal{N} \rceil/2$, is roughly negligible on a large L

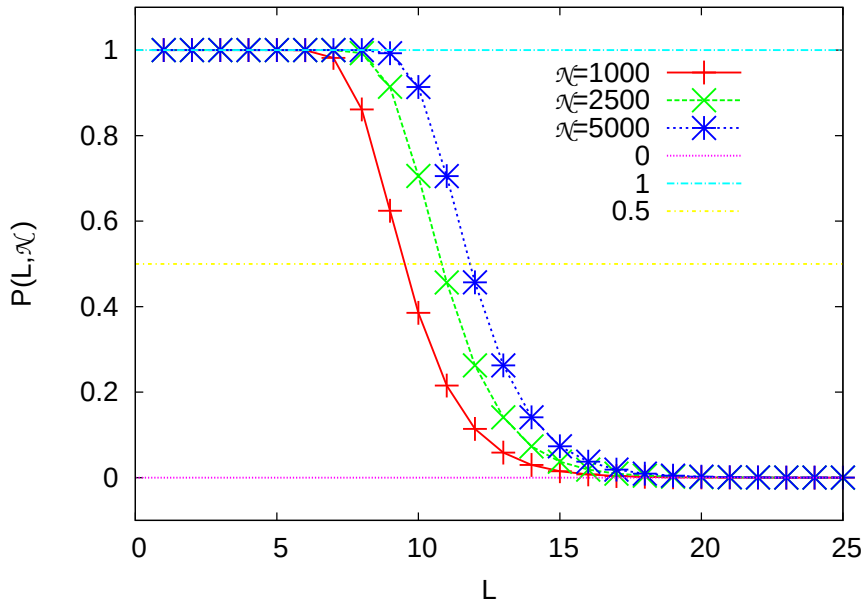

Figure 2.16: **Numerical evaluation of L^* for the considered PS model.** Plots of the result of the calculation of $P(L,\mathcal{N})$ for $\mathcal{N}$=1000, 2500 and 5000. One gets correspondingly $L^* \simeq 10-11$.

range, up to small $L \ll \mathcal{N}/2$ values, for large $\mathcal{N}$ values. Furthermore, in the interesting L-range in which one finds the rapid decrease of $P(L,\mathcal{N})$, in particular for $0.3 \leq P(L,\mathcal{N}) \leq 0.7$, the difference between the correct $P(L,\mathcal{N})$ and the closed formula (2.55) is both relatively small and roughly constant.

On such bases, it should be possible to obtain a refined evaluation from below of $\mathcal{N}^*$ also in the case of L^* of the order of more than a hundred of base pairs. In fact, we already obtained a relatively small $L^* \sim 50$ for $\mathcal{N}^*$ as large as 10^7, hence, in order to evaluate the crossover chain length for the PS model studied in [38, 39], one expects to have to calculate $P(L,\mathcal{N})$ for $\mathcal{N}^* \gg 10^7$ values. In this last case, whereas the practical calculation of this quantity would appear impossible by using the L-step Fibonacci number recursion relations (because of the numbers in these sequences being very rapidly increasing, as it is well known, too), a much better approximation than (2.33) should be made possible by using (2.55) and our just discussed preliminary observations.

Chapter 3

Applying the SPDERG approach to Michaelis-Menten kinetics

The original results that we are going to present in the second part of this thesis have been obtained in collaboration with A.M. Bersani and E. Bersani, and they appeared in [40]. As anticipated, they concern a rigorous way of applying in particular to Michaelis-Menten (MM) kinetics [28] an approach proposed in [26, 27] as a possible alternative to the standard Perturbation Method to deal with Singularly Perturbed Differential Equation, that is based on a Renormalization Group approach (SPDERG).

Therefore, we start by a quite general discussion that should clarify some basic RG ideas, in such a way both to underline the connection, from the theoretical point of view, with the first part of the thesis and to make the SPDERG approach itself more intuitively understandable.

Finally, we introduce the SPDERG approach and we present our results.

3.1 The Renormalization Group approach

The analytical results obtained by RG techniques that we recalled in the previous Chapter (see for instance [14, 18, 19]), in order to sketch the background of our work on DNA denaturation, were examples of applications of this approach that aimed to understand the behaviour near the critical point of Condensed Matter systems. In fact, though historically the Renormalization Group was introduced within Quantum Field Theories, the basis for the formulation of the same principles in a way appropriate for studies in Condensed Matter (Statistical Field Theory) dates back to the beginning of the Seventies of the last century, when appeared in particular the work by Wilson [76], at the origin of his Nobel prize. In the meanwhile, the RG has been both largely and successfully used in these research fields, in which the approach can be expressed in very rigorous terms [8].

On the other hand, the basic property required for the application of RG techniques is merely some kind of scale invariance, that is encountered in the study of many other systems, like in Physics, in Biomathematics and in Applied Mathematics. In the present work, clearly, we are interested to the application of these techniques to Dynamical Systems, a research field in which they can be generally useful for understanding the asymptotic system behaviour and, as originally proposed in particular by Chen, Goldenfeld and Oono [26], for removing large time divergences and for finding out (usually at a given order in an appropriate expansion parameter ε) the solution of singularly perturbed differential equations.

In fact, in the literature one finds a large number of different strategies for applying RG techniques, depending on the kind of study one is interested in. Anyway, let us start by recalling the basis of the standard RG procedure [9, 41], that consists of two steps: a *coarse graining* one, in which the averages on a subgroup of the system's degrees of freedom are performed, and a *rescaling* one, in which the

length unity is redefined, *i.e.*, multiplied by an appropriate scaling factor. Then, the RG equations relate the renormalized parameters $\{\theta_i(1); i = 1, \dots n_p\}$ to the original ones $\{\theta_i(0); i = 1, \dots n_p\}$. By repeating an enough large number of times, n^*, this two-step procedure, because of the hypothesized scale invariance, the set of the renormalized parameters reaches a fixed point, *i.e.*, one expects that, after n^* repetitions, $\{\theta_i(n^* + 1)\} = \{\theta_i(n^*)\} = \{\theta_i^*\}$ or that, in any event, the fixed point is reached asymptotically, for $n^* \to \infty$.

A simple example [9, 41], that also reminds, from a different point of view, the polymer subject of the previous Chapter, is the one of the Random Walk (RW). In order to clarify the differences, we notice that here one considers once more a (single) RW in D-dimension, that has the origin as starting point, but one studies its behaviour as function of time t. Most importantly, whereas in the previous case the "polymeric" configuration of length $\mathcal{N}$ could be considered as made by the end point path after $\mathcal{N}$ time steps, here we look in particular just at the end point position with respect to the origin itself, and we are interested in calculating the probability distribution of its average total displacement as a function of t.

Without loss of generality, one can assume from the beginning to be in the continuous space $\Re^D$ rather than on a hyper-cubic lattice (in any event, one could arrive to the continuous space just by coarse graining and rescaling, as in other cases). Then, by definition, the direction of the RW displacement at each t-step is an uniformly distributed random variable and one can further assume that each displacement, $\vec{r} \in \Re^D$, is a random variable whose modulus, $r \equiv |\vec{r}|$, obeys to a Gaussian distribution of variance σ_r:

$$\mathcal{P}(r) = \frac{1}{\sqrt{2\pi\sigma_r^2}} e^{-r^2/(2\sigma_r^2)}. \tag{3.1}$$

Then, after n_t time steps, with the RW that makes a displacement $\vec{r}_i$ $(i = 1, \dots, n_t)$ at each t-step, the RG coarse graining corresponds to study the total displacement variable:

$$\vec{r}_{tot} = \sum_{i=1}^{n_t} \vec{r}_i, \tag{3.2}$$

whose modulus, $r_{tot} \equiv |\vec{r}_{tot}|$ turns out to be distributed according to:

$$\mathcal{P}(r_{tot}) = \frac{1}{\sqrt{2\pi n_t\sigma_r^2}} e^{-r_{tot}^2/(2n_t\sigma_r^2)}. \tag{3.3}$$

Hence, the coarse graining gives the same distribution with $\sigma_r \to \sqrt{n_t}\sigma_r$. Correspondingly, the RG rescaling is to be made by taking:

$$\vec{r}_i \to \sqrt{n_t}\vec{r}_i. \tag{3.4}$$

With this choice, the renormalized distribution probability (*i.e.*, the one for r_{tot}) is exactly the same as the starting distribution probability (3.1), with the same variance, and the fixed point is reached at the first application of the RG procedure. In fact, here the system is already at its fixed, or *critical*, point. In this simple example, it is also evident from the RG application that, since $\sigma_{r_{tot}} = \sqrt{n_t}\sigma_r$, whereas, on the other hand, the RG equation for σ_r is just the identity, the average distance of the RW from the origin increases as $\sim (n_t)^{1/2}\sigma_r \propto (n_t)^{1/2}$. In other words, after a time t the RW has moved away from the origin of a distance in average $\propto t^{1/2}$ or, equivalently (by dimensional analysis), the RW correlation length exponent is $\nu = 1/2$, a well known result.

Usually, one faces more complicated situation. These are characterized, first of all, by the presence of divergent terms that need to be eliminated. For instance, one encounters: i) divergences in the ultra-violet wave length regime in Quantum Field Theory; ii) divergences in the infra-red wave length regime in Statistical Field Theory; iii) divergences at large times (the secular terms) when applying the RG to Dynamical Systems. Very roughly and generally speaking, in order to deal with such a kind of problems one usually introduces some *ad hoc* parameters (for instance, an appropriate cut-off in the ultra-violet or in the infra-red, respectively). Correspondingly, the RG scaling equations

are to be solved in such a way that one could also end up with a solution independent from these parameters, since they were not present in the initial theory, hence they have no physical meaning. These are the key ingredients on the basis of which one can calculate the renormalized mass of the electron, to mention another well known result.

As we are going to discuss item by item, by recalling the discussion in [26, 27] in the case of boundary layer problems, as the one posed by MM kinetics, these are the key ingredients of the SPDERG approach, too.

3.2 Michaelis-Menten kinetics

We rewrite here the enzymatic reaction at the basis of MM kinetics, that we anticipated in the introduction (see (1.1)):

$$\mathcal{E} + \mathcal{S} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathcal{C} \xrightarrow{k_2} \mathcal{E} + \mathcal{P}. \tag{3.5}$$

Here, $\mathcal{E}$, $\mathcal{S}$, $\mathcal{C}$ and $\mathcal{P}$ label the Enzyme, the Substrate, the Complex and the Product, respectively, whereas k_1, k_{-1} and k_2 are the reaction kinetic constants. A graphic scheme of this kind of reaction is presented in [Fig. 3.1].

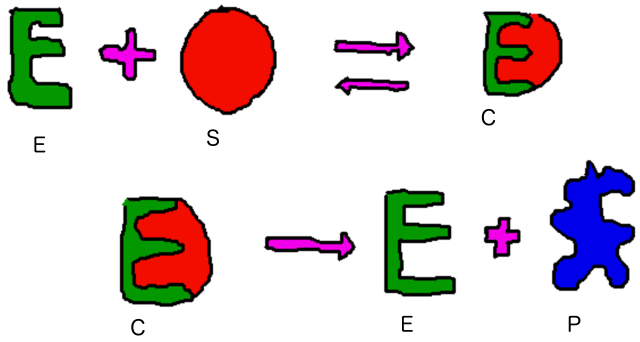

Figure 3.1: **Scheme of the fundamental reaction in MM kinetics.** As reported in (3.5) in the text, here the Enzyme and the Substrate transform into the Complex by a reversible reaction, whereas, at the same time, the Complex transform into the Enzyme and the Product by an irreversible reaction.

In fact, MM kinetics [28] is a well known example in Biomathematics [1, 2, 3, 4] of a system of ODEs characterized by two definitely different time scales. At the very beginning the complex rapidly reaches a pseudo-equilibrium with the substrate in a time interval whose length can be taken as the first time scale to be considered. The second part of the reaction, that is important for the experimental observations, happens on a time scale that can be as longer as several orders of magnitude than the first one [77, 30].

Therefore, from the theoretical point of view [29, 1, 2, 10], MM kinetics is an example of boundary layer problem. In the literature there are a lot of papers aiming at approximating the solutions of the system. In the *standard* QSSA (sQSSA) [28, 29, 1, 2], the independent variables are chosen to be the substrate and the complex, with the complex evolution that depends algebraically on the substrate. It is the routinely taken starting point for investigating the system's dynamics. Nevertheless, different starting points are taken into consideration in the literature [29, 1].

As advanced, here we are interested in testing the correctness, in MM kinetics, of the SPDERG approach proposed by Chen, Goldenfeld and Oono in [26, 27], particularly as an alternative to the

perturbation expansion. In order to clarify similarities and differences between these two analytical ways of approximating the correct solutions, we focus here on the case of the perturbation expansion beyond the sQSSA, since the sQSSA appears the simplest starting point from the analytical point of view [29].

Indeed, in the standard method, the solutions are approximated just by the perturbation expansion, in an appropriate parameter ε, of both the inner and the outer components of the two chosen independent variables, *i.e.*, of the solutions of systems of two ODEs with regular and singular perturbations, respectively, with the further imposition, at each order in ε, of the appropriate matching conditions. Within this framework, both the sQSSA [29, 1, 2, 31] and the *total* QSSA (tQSSA) [29, 78, 79, 80, 81, 82, 83, 84] represent the 0th order terms of the outer solutions. From the mathematical point of view, the correctness of applying the perturbation expansion to MM kinetics can be justified on the basis of the Tikhonov theorem [1, 31], as well as on the basis of different theoretical results [2], in both of the cases by assuming a small enough expansion parameter ε.

3.2.1 The sQSSA

The sQSSA represents a milestone in the mathematical modelling of enzymatic reactions [28, 29, 1, 2, 3, 4, 85, 86, 87, 88]. Here we just remind that the original paper by Michaelis and Menten dates back to more than one century ago [28], that the idea was already present in the previous papers by Henri [85, 86, 87], and that the approach was further developed in particular by Briggs and Haldane [88]. Schematically [29, 1], one is modelling the reaction (3.5) that is reversible in the first part and irreversible in the second one, with associated kinetic constants k_1, k_{-1}, and k_2.

When introducing the concentrations e, s, c, and p, respectively, by using the mass action law, one arrives to describe the process by means of a system of four 1st order ODEs. Then, within the standard framework [29, 1], one starts by using the conservation law $e + c = e_0 + c_0$, that implies that the enzyme concentration, e, does only depend on the complex one, c. Moreover, one observes that the product concentration, p, can be obtained from the complex concentration, c, by integrating (equivalently, one can use the other conservation law, $s + c + p = s_0 + c_0 + p_0$). Finally, one assumes that the concentrations of the complex and of the product are zero at the beginning, as in typical experiments (*i.e.*, $c_0 = p_0 = 0$).

Hence, one ends up with the well known system of two 1st order ODEs (that we write in the same form as in [1]), that are to be obeyed by the variables s and c (with ICs $s(0) = s_0$ and $c(0) = 0$):

$$\begin{cases} \dot{s}(t) &= k_1[s(t) + K_D]\left[c(t) - \dfrac{e_0 s(t)}{s(t) + K_D}\right] \\ \dot{c}(t) &= -k_1[s(t) + K_M]\left[c(t) - \dfrac{e_0 s(t)}{s(t) + K_M}\right], \end{cases} \tag{3.6}$$

where the dot means the time derivative.

Here $K_D = k_{-1}/k_1$ is the so-called dissociation constant, whereas $K_M = (k_{-1} + k_2)/k_1$ is the parameter that is generally known as Michaelis constant. It can be further noticed that $K_M - K_D = k_2/k_1 = K$ is the Van Slyke-Cullen constant [89]. Though the original kinetic constants, k_1, k_{-1} and k_2, are the key physical parameters for the studied system, only K_M, together with $V_{max} = k_2 e_0$, turns out to be experimentally measurable, in particular on the basis of the sQSSA.

The kinetic constants that we choose are (as in [29]):

$$k_1 = 1\mu M^{-1}s^{-1};\;\; k_{-1} = 4s^{-1};\;\; k_2 = 1s^{-1}; \tag{3.7}$$

As we clarify with the following discussion on the solution behaviours, both this choice and the one of the two sets of ICs, that we are going to introduce, correspond to cases in which the sQSSA is a worse approximation than for different possible ones. Indeed, we make these choices just because

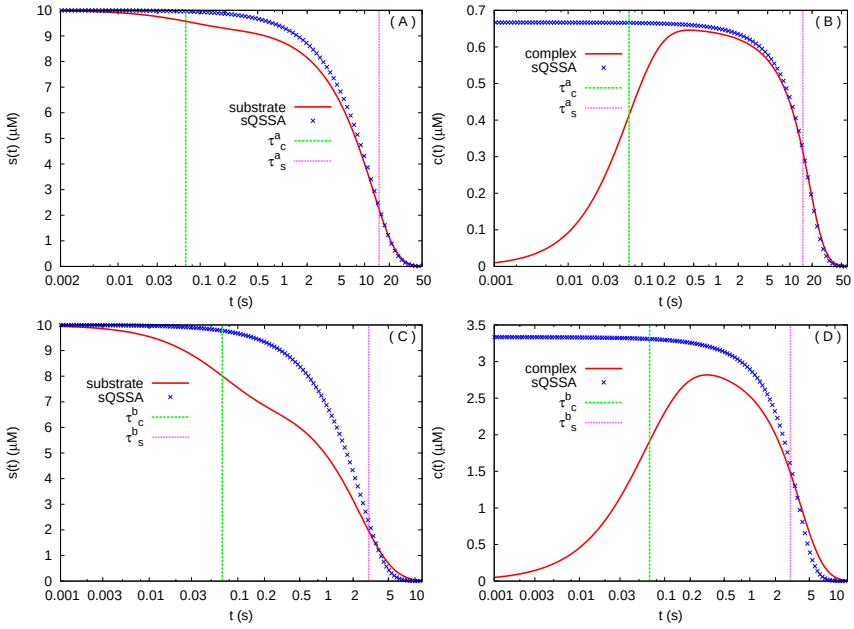

Figure 3.2: **The numerical solutions, for the two considered sets of ICs, and its comparison with the sQSSA ones.** In A) and in C) we present the behaviour of the concentrations of the substrate $s(t)$, whereas in B) and in D) we present the one of the concentrations of the complex $c(t)$, solutions of Eqs. (3.6), for the a and b sets of ICs given by (3.8), respectively. Notice that the time is in logarithmic scale. We plot the corresponding rough evaluations of the two different time scales involved, too, with τ_s describing the substrate decay time and τ_c the complex saturation time (see the text for details). Moreover, we present in advance the analytical solutions computed from the sQSSA ones (by using a standard numerical approximation to the Lambert function) as given by (3.12) and (3.14), whose derivation will be recalled in the following of the present Section (from this point of view, notice that in A) and B) we are in the case in which the standard expansion parameter is $\varepsilon = \varepsilon^a = 0.1$, whereas in C) and D) we are in the one with $\varepsilon = \varepsilon^b = 0.5$).

they are appropriate to the present analysis. In particular, these kinetic constants give $K_D = 4\mu M$, $K_M = 5\mu M$ and $K = 1\mu M$, whereas the two sets of ICs that we consider are:

$$e_0^a = 1\mu M; \quad e_0^b = 5\mu M; \quad s_0^a = s_0^b = 10\mu M; \tag{3.8}$$

In detail, we are taking different values of the initial enzyme concentration, e_0, with $e_0^b > e_0^a$, by labeling a and b the two corresponding sets.

We plot in [Fig. 3.2] the solutions of Eqs. (3.6), obtained by numerically integrating the system, for these two sets of ICs, respectively. In the same figure, we compare the solutions with their sQSSA, by enlightening the fact that MM kinetics cannot be confused with the MM approximation (or sQSSA), that is expected to be valid only in the second, pseudo-equilibrium phase. The expected inadequacy of the sQSSA in the transient phase is enhanced by the use of the logarithmic scale for the time variable, which enlarges the transient phase. On the other hand, the logarithmic scale shrinks the intervals for large times, not allowing us to appreciate the bad approximation of the sQSSA for large times, when $\varepsilon = 0.5$.

The logarithmic scale in the figures makes evident the typical presence of two definitely different time scales, with the concentration of the complex, c, that evolves very rapidly at the beginning, whereas it turns out to be in a *quasi-steady-state* or *pseudo-equilibrium* in the second part, of definitely longer duration [90]. Indeed, in logarithmic scale, the presence of a *plateau*, in which the complex is really roughly constant, is well more evident in case a, whereas in case b it displays a more or less

symmetric behaviour around the maximum. Actually, as we are going to better discuss, this last choice of the ICs corresponds to a particularly unfavourable situation for applying the sQSSA.

In fact, in the sQSSA, one focuses on the second part, by assuming that the complex depends algebraically on the substrate. Indeed, this is equivalent to take $\dot{c} \sim 0$ in Eqs. (3.6). Therefore, from the mathematical point of view, the sQSSA appears quite a coarse approximation, that has been correspondingly largely criticized in the literature. In particular, we refer the reader to the detailed review [29], where the experimental cases, in which the sQSSA fails, are resumed, too, and to the work by Heineken, Tsuchiya and Aris [31], where the perturbation expansion beyond the sQSSA, with our choice of the ε expansion parameter, is originally presented. In any event, the sQSSA is clearly the more correct on the whole time window the more the time scale corresponding to the rapid transient phase of the complex, τ_c, is short with respect to the time scale τ_s, that rules the slow decay of the substrate.

To continue, let us note that one can argue about different time scale evaluations (see for instance the recent work in [91]). Here, we consider a simple approach in which the two most relevant time scales to the system's dynamics are just the initial complex saturation time, τ_c, and the final substrate decay time, τ_s. In fact, we are considering the evaluations of τ_s and τ_c given in [1], that indeed coincide with the ones obtained by quite detailed reasonings in [92]: on the one hand, from the assumption $\dot{c} \sim -(c - c_{eq})/\tau_c$, with $s \sim s_0$, one gets $\tau_c = 1/(k_1(s_0 + K_M))$; on the other hand, within the sQSSA, one has $\dot{s} \sim -k_2 e_0 s/(s + K_M)$, hence $\dot{s} \sim -s/\tau_s$, with $\tau_s = (s_0 + K_M)/(k_2 e_0)$, if one moreover assumes a final exponential substrate decay. These evaluations give the same values for the complex saturation times, $\tau_c^a = \tau_c^b = 0.0667s$, in our case, for the a and b sets of ICs, respectively, whereas the substrate decay times are $\tau_s^a = 15s$ and $\tau_s^b = 3s$, respectively, *i.e.*, τ_s is larger for the a set of ICs than for the b one.

Let us now look more in detail at the behaviours displayed by the present numerical solutions of the original problem, given by Eqs. (3.6), when plotted in logarithmic scale. Indeed, in the case of the substrate, one observes, both in [Fig. 3.2A] and in [Fig. 3.2B], the presence of three inflection points. Importantly, this feature appears to depend on the present choices of the kinetic constants and of the ICs. We numerically checked that, for instance, in the case $k_1 = 1\mu M^{-1}s^{-1}$, $k_{-1} = 10s^{-1}$ and $k_2 = 20s^{-1}$, with the a set of ICs, as well as with the present kinetic constants, but by taking a definitely smaller value of $e_0 = 0.1\mu M$, one finds a (not shown) simpler behaviour than here, with only one inflection point in the curve of the substrate.

These observations confirm that we are considering situations in which the correct solutions are difficult to be analytically approximated, at least by the perturbation expansion that goes beyond the sQSSA, just since one has to reproduce a behaviour of the substrate curve that is particularly complicated. Correspondingly, our choice of the kinetic constants and of the two sets of ICs is justified.

In the light of the same observations, it is also even more interesting that, as can be seen from [Fig. 3.2], in logarithmic scale, the obtained estimation of τ_s captures quite accurately the last (*i.e.*, the third) inflection point in the curve of the substrate, that therefore appears, correctly, interpretable as the substrate decay time. On the other hand, the obtained estimation for τ_c captures quite accurately the first inflection point in the curve of the complex, that therefore appears, correctly, interpretable as the complex saturation time, too. In fact, the position of the third inflection point in the curve of the substrate roughly coincides with the one of the second inflection point in the curve of the complex. Correspondingly, both of the inflection points in the curve of the complex are captured by τ_c and τ_s, respectively (this is more evident in the case of the a set of ICs). Moreover, the complex saturation time τ_c turns out to be quite close to the first inflection point in the curve of the substrate.

With the aim of deepening the analysis, we are led to the problem of making dimensionless the system given by Eqs. (3.6). This involves the choice of the ε variable as the one giving the sQSSA condition for $\varepsilon = 0$. In fact, such a variable is clearly also the possible expansion parameter for going beyond the sQSSA. Here, within the two known possible schemes for making dimensionless

the equations, and the two corresponding different choices of the ε variable [29, 1, 92], in the sQSSA case, we limit ourselves to study the one that is more largely considered in the literature [88, 29, 1, 3, 4, 2, 31], *i.e.*, $\varepsilon = e_0/s_0$.

In detail (again following in particular [1], in the notation, too), one introduces the dimensionless parameters $m = K_D/s_0$, $M = K_M/s_0$, and one scales the time of a factor δ, $t \to \delta t$ with $\delta = k_1 e_0$. The substrate concentration is made dimensionless by taking $\tilde{s}(t) = s(t)/s_0$. Moreover, in the presently considered scheme, the complex concentration is made dimensionless by taking $\tilde{c}(t) = c(t)/e_0$ (that is a possible correct choice since, thanks to the first of the recalled conservation laws, $c(t) \leq e_0$).

Correspondingly, one obtains the (singular with respect to ε) system of ODEs:

$$\left\{ \begin{array}{rcl} \dot{\tilde{s}}^{out}(t) & = & [\tilde{s}^{out}(t) + m] \left[\tilde{c}^{out}(t) - \dfrac{\tilde{s}^{out}(t)}{\tilde{s}^{out}(t) + m} \right] \\ \varepsilon \dot{\tilde{c}}^{out}(t) & = & -[\tilde{s}^{out}(t) + M] \left[\tilde{c}^{out}(t) - \dfrac{\tilde{s}^{out}(t)}{\tilde{s}^{out}(t) + M} \right]. \end{array} \right. \tag{3.9}$$

Here, the label *out* refers to the fact that these are, in fact, the ODEs that capture the long time behaviours, *i.e.*, the ones to be obeyed by the *outer* solutions. With the present kinetic constant choice (3.7), we have $m = 0.4$ and $M = 0.5$, respectively. On the other hand, the two sets of ICs given by (3.8) yield two different values for δ and for the expansion parameter ε:

$$\begin{array}{rclrcl} \delta^a & = & 1 s^{-1}; & \delta^b & = & 5 s^{-1}; \\ \varepsilon^a & = & 0.1; & \varepsilon^b & = & 0.5; \end{array} \tag{3.10}$$

Thus, though the basic condition $\varepsilon < 1$ is verified in both of the cases, we are in the situation $\varepsilon^b > \varepsilon^a$, *i.e.*, the one expected to be appropriate to highlight differences between the approximations.

The system (3.9) should make evident that the sQSSA can be rigorously interpreted as the 0th order term of an asymptotic expansion in ε of this example of singular perturbation [29, 1, 2, 10], in which the original system is reduced to one ODE and one algebraic relation. For $\varepsilon = 0$, one has:

$$\left\{ \begin{array}{rcl} \dot{\tilde{s}}_0^{out}(t) & = & -\dfrac{M - m}{\tilde{s}_0^{out}(t) + M} \tilde{s}_0^{out}(t) \\ \tilde{c}_0^{out}(t) & = & \dfrac{\tilde{s}_0^{out}(t)}{\tilde{s}_0^{out}(t) + M}. \end{array} \right. \tag{3.11}$$

It is to be noted that, in the numerator of the ODE to be obeyed by $\tilde{s}_0^{out}(t)$, the quantity $M - m$ is just the dimensionless Van Slyke-Cullen constant [89], K/s_0. Here, the system is to be considered together with the substrate IC, $\tilde{s}_0^{out}(0) = 1$, in fact implying that the complex initial value is automatically fixed to $\tilde{c}_0^{out}(0) = 1/(1 + M)$.

The ODE for the dimensionless substrate concentration can be solved explicitly [29, 93], by means of the Lambert function $\omega(x)$ [94], that verifies the equation $\omega(x) e^{\omega(x)} = x$:

$$\tilde{s}_0^{out}(t) = M \omega(e^{-(M - m)t/M + 1/M}/M). \tag{3.12}$$

This solution does also satisfy the IC, $\tilde{s}_0^{out}(0) = 1$, since:

$$M \omega(e^{1/M}/M) = M \omega\left[\omega^{-1}(1/M)\right]. \tag{3.13}$$

Correspondingly, one gets:

$$\tilde{c}_0^{out}(t) = \frac{\omega(e^{-(M - m)t/M + 1/M}/M)}{\omega(e^{-(M - m)t/M + 1/M}/M) + 1}, \tag{3.14}$$

with, in particular, $\tilde{c}_0^{out}(0) = 1/(1 + M)$, as expected.

Hence, by using a standard numerical approximation to the Lambert function, we obtain the behaviours of $\tilde{s}_0^{out}(t)$ and $\tilde{c}_0^{out}(t)$, *i.e.*, the behaviours of the substrate and complex concentrations within the sQSSA. We already presented in [Fig. 3.2] our results for the two considered sets of ICs, in comparison with the numerical solution of the original problem, given by Eqs. (3.6).

We note, qualitatively, that the sQSSA gives definitely worse results for the b set of ICs ([Fig. 3.2C] and [Fig. 3.2D]), as predictable on the basis that $\varepsilon^b > \varepsilon^a$. In both cases, in these figures the logarithmic scale highlights that the complex very short transient time behaviour is not captured at all. Indeed, this is the well known failure of the sQSSA (that is shared by the tQSSA, too) that, as we already recalled, is quite largely criticized in the literature [29, 31]. From the mathematical point of view, it is to be stressed again that, whereas one can solve Eqs. (3.11) by choosing $\tilde{s}_0^{out}(0) = 1$, the original IC of the complex cannot *a priori* be satisfied within the present context. Thus, the sQSSA is unable to predict the initial increase from zero of this quantity.

Furthermore, the presence of more than one inflection point in logarithmic scale in the curve of the substrate (hence the short time qualitative behaviour of this quantity) is not reproduced at all. In fact, also the maximum reached by the complex during its evolution is lower than its initial value within the sQSSA in both of the considered cases. This last feature is not observed in the other previously mentioned cases, that display simpler solution behaviours than the present ones, as we numerically checked. On this basis, one can indeed hypothesize that the various effects are related, though we left a more rigorous analysis to future studies. Presumably for the same reason, when taking $\varepsilon = \varepsilon^b = 0.5$ in the case b, it is naked-eye evident that the sQSSA fails in reproducing the long time behaviour, too (see [Fig. 3.2C] and [Fig. 3.2D]). Thus, our choices turn out to be even more appropriate than expected to the present analysis.

3.3 Standard perturbation method

3.3.1 Known results beyond the sQSSA

The standard perturbation expansion method, also in the case of MM kinetics [29, 1, 2, 3, 4], is based on mathematical results for systems of ODEs with both singular and regular perturbations [10]. Within this framework, the system given by Eqs. (3.9) needs to be considered together with the system that one obtains with the transformation $t \to \tau = t/\varepsilon$, *i.e.*, in the opposite limit of short times, that is:

$$\left\{ \begin{array}{rcl} \dot{\tilde{s}}^{in}(\tau) & = & \varepsilon[\tilde{s}^{in}(\tau) + m]\left[\tilde{c}^{in}(\tau) - \dfrac{\tilde{s}^{in}(\tau)}{\tilde{s}^{in}(\tau) + m}\right] \\ \dot{\tilde{c}}^{in}(\tau) & = & -[\tilde{s}^{in}(\tau) + M]\left[\tilde{c}^{in}(\tau) - \dfrac{\tilde{s}^{in}(\tau)}{\tilde{s}^{in}(\tau) + M}\right], \end{array} \right. \tag{3.15}$$

to be instead obeyed by the *inner* solutions.

Correspondingly [29, 1, 2, 3, 4, 10], one looks for solutions in the form:

$$\left\{ \begin{array}{rclcrcl} \tilde{s}^{in}(\tau) & = & \displaystyle\sum_{i=0}^{\infty} \tilde{s}_i^{in}(\tau)\varepsilon^i; & & \tilde{c}^{in}(\tau) & = & \displaystyle\sum_{i=0}^{\infty} \tilde{c}_i^{in}(\tau)\varepsilon^i; \\ \tilde{s}^{out}(t) & = & \displaystyle\sum_{i=0}^{\infty} \tilde{s}_i^{out}(t)\varepsilon^i; & & \tilde{c}^{out}(t) & = & \displaystyle\sum_{i=0}^{\infty} \tilde{c}_i^{out}(t)\varepsilon^i; \end{array} \right. \tag{3.16}$$

by requiring that $\left\{\tilde{s}_i^{in}(\tau), \tilde{c}_i^{in}(\tau)\right\}$ satisfy the system given by Eqs. (3.15), in the case of the inner solutions, and that $\left\{\tilde{s}_i^{out}(t), \tilde{c}_i^{out}(t)\right\}$ satisfy the system given by Eqs. (3.9), in the case of the outer solutions, at each order in ε. One then imposes the appropriate matching conditions and one takes, as uniform approximations to the correct solutions at a given order in ε, the sum of the inner and of the outer solutions at that order minus the common terms [29, 2].

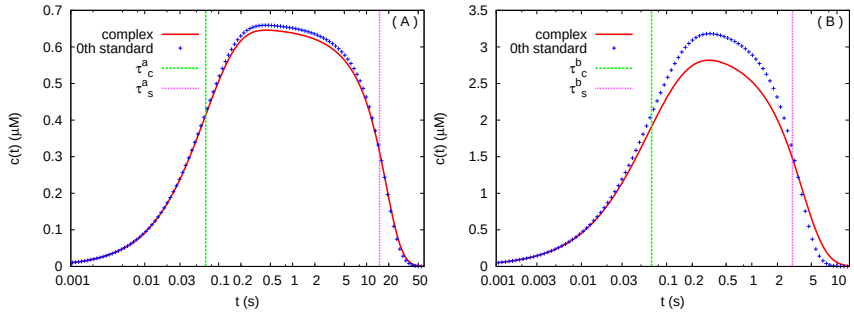

Figure 3.3: **Comparison between the numerical solutions and the 0th order perturbation expansion uniform approximation, for the two considered sets of ICs, for the complex.** In A) we present the behaviour of the concentration of the complex $c(t)$ for the a set of ICs given by (3.8), whereas in B) we present the one for the b set. Hence, it is $\varepsilon = \varepsilon^a = 0.1$ in the first case and $\varepsilon = \varepsilon^b = 0.5$ in the second one. We plot both the numerical solutions of Eqs. (3.6), already shown in the previous figure, and the analytical solutions computed from the dimensionless 0th order perturbation expansion uniform approximation (by using a standard numerical approximation to the Lambert function) as given by (3.19). We finally plot our corresponding rough evaluations of the two different time scales involved, too, with τ_s describing the substrate decay time and τ_c the complex saturation time. Notice that the time is in logarithmic scale.

The whole procedure is therefore a standard perturbation expansion method [10]. We, moreover, remind that the need for imposing appropriate matching conditions is due to the lack of ICs for the outer solutions; they can be determined from the behaviours of the inner ones in the large τ limit, that have to correspond to their behaviours in the small t limit (whereas the inner solutions are the ones that satisfy the original ICs of the problem) [29, 1, 2, 10].

Let us, first of all, recall the 0th order inner solutions, obtained by setting $\varepsilon = 0$ in Eqs. (3.15). These solutions are [29, 1]:

$$\begin{cases} \tilde{s}_0^{in}(\tau) &= 1 \\ \tilde{c}_0^{in}(\tau) &= \dfrac{1}{1+M}\left[1 - e^{-(1+M)\tau}\right]. \end{cases} \tag{3.17}$$

Let us continue by recalling also the 0th order outer solution of the system given by Eqs. (3.11), with IC $\tilde{s}_0^{out}(0) = \tilde{s}_0^{out*}$, where $\tilde{s}_0^{out*}$ needs to be determined by imposing the matching condition on the substrate. In the relatively simple 0th order case, one expects that the substrate solution verifies $\tilde{s}_0^{out*} = 1$. In fact, with this choice, one has:

$$\begin{cases} \lim_{\tau\to\infty} \tilde{s}_0^{in}(\tau) &= 1 &= \lim_{t\to 0} \tilde{s}_0^{out}(t) \\ \lim_{\tau\to\infty} \tilde{c}_0^{in}(\tau) &= \dfrac{1}{1+M} &= \lim_{t\to 0} \tilde{c}_0^{out}(t). \end{cases} \tag{3.18}$$

Hence, not only the substrate matching condition turns out to be satisfied, but, consistently [2], this condition (the only one to be freely enforceable) satisfies the complex matching condition, too.

In conclusion, the 0th order perturbation expansion uniform approximations (that we label u), are given by [29, 1]:

$$\begin{cases} \tilde{s}_0^u(t) = \tilde{s}_0^{in}(t/\varepsilon) + \tilde{s}_0^{out}(t) - 1 + O(\varepsilon) \\ \tilde{c}_0^u(t) = \tilde{c}_0^{in}(t/\varepsilon) + \tilde{c}_0^{out}(t) - \dfrac{1}{1+M} + O(\varepsilon). \end{cases} \tag{3.19}$$

We plot in [Fig. 3.3] our results on the complex concentrations (the uniform approximation to the substrate being coincident with the sQSSA one), for the two considered sets of ICs, in comparison with the numerical solutions of the original problem, given by Eqs. (3.6) (the same curves as in [Fig.

3.2]). The figures show that this approximation already captures the most characteristic features of the whole system's dynamics and, in particular, the rapid initial increase of the complex.

Nevertheless, the approximation over-evaluates the correct maximum values of the complex, as it is definitely more evident in case b, that corresponds to the higher value of the expansion parameter, $\varepsilon = \varepsilon^b = 0.5$, of the two sets of ICs that we consider ([Fig. 3.3B]). On the other hand, as we already outlined, one can notice in the curves of the substrate ([Fig. 3.2A] and [Fig. 3.2C]) the sQSSA failure in capturing the presence of the first two inflection points (in logarithmic scale). The presently recalled results make clear that this failure is shared by the 0th order perturbation expansion uniform approximation, too.

Let us, finally, recall the calculation of the 1st order contribution [2, 31]. The 1st order inner solutions solve the system:

$$\begin{cases} \dot{\tilde{s}}_1^{in}(\tau) &= (1+m)\left[\tilde{c}_0^{in}(\tau) - \dfrac{1}{1+m}\right] \\ \dot{\tilde{c}}_1^{in}(\tau) &= -(1+M)\tilde{c}_1^{in}(\tau) - \left[\tilde{c}_0^{in}(\tau) - 1\right]\tilde{s}_1^{in}(\tau). \end{cases} \tag{3.20}$$

Here, we already used $\tilde{s}_0^{in}(\tau) = 1$, thus the ICs are $\tilde{s}_1^{in}(0) = \tilde{c}_1^{in}(0) = 0$. In fact, one has in particular to solve an ODE for $\tilde{c}_1^{in}$ in the form $\dot{y}(\tau) = -(1+M)y(\tau) + f(\tau)$, whose solution is $y(\tau) = a(\tau)e^{-(1+M)\tau}$, with $a(\tau) = \int_0^\tau f(z)e^{(1+M)z}dz$. One obtains [2, 31]:

$$\begin{cases} \tilde{s}_1^{in}(\tau) &= -\dfrac{M-m}{1+M}\tau - \dfrac{1+m}{(1+M)^2}\left[1 - e^{-(1+M)\tau}\right] \\ \tilde{c}_1^{in}(\tau) &= -\dfrac{M(M-m)}{(1+M)^3}\tau - \dfrac{M(1+2m-M)}{(1+M)^4}\left[1 - e^{-(1+M)\tau}\right] + \\ &- \left[\dfrac{(1-M)(1+m)}{(1+M)^3}\tau + \dfrac{M-m}{(1+M)^2}\dfrac{\tau^2}{2}\right]e^{-(1+M)\tau} + \\ &+ \dfrac{(1+m)}{(1+M)^4}e^{-(1+M)\tau}\left[1 - e^{-(1+M)\tau}\right]. \end{cases} \tag{3.21}$$

On the other hand, the 1st order outer solutions are more complicated to be obtained than the inner ones, since they solve the system (that consists once again of one ODE and of one algebraic relation):

$$\begin{cases} \dot{\tilde{s}}_1^{out}(t) &= \dfrac{M(M-m)}{\left[\tilde{s}_0^{out}(t) + M\right]^4}\tilde{s}_0^{out}(t)\left[\tilde{s}_0^{out}(t) + m\right] - \dfrac{M(M-m)}{\left[\tilde{s}_0^{out}(t) + M\right]^2}\tilde{s}_1^{out}(t) \\ \tilde{c}_1^{out}(t) &= \dfrac{M(M-m)}{\left[\tilde{s}_0^{out}(t) + M\right]^4} + \dfrac{M}{\left[\tilde{s}_0^{out}(t) + M\right]^2}\tilde{s}_1^{out}(t), \end{cases} \tag{3.22}$$

in which $\tilde{s}_0^{out}(t)$ is reported, in terms of the Lambert function, in (3.12). One can verify that the solution for $\tilde{s}_1^{out}(t)$ is given by [2]:

$$\tilde{s}_1^{out}(t) = \frac{s_0^{out}(t)}{\tilde{s}_0^{out}(t) + M}\left\{\frac{m}{M}\log\left[\frac{\tilde{s}_0^{out}(t) + M}{(1+M)\tilde{s}_0^{out}(t)}\right] - \frac{\tilde{s}_0^{out}(t) + m}{\tilde{s}_0^{out}(t) + M}\right\}. \tag{3.23}$$

Correspondingly, the 1st order outer solution for the complex is:

$$\tilde{c}_1^{out}(t) = \frac{\tilde{s}_0^{out}(t)}{\left[\tilde{s}_0^{out}(t) + M\right]^3}\left\{m\log\left[\frac{\tilde{s}_0^{out}(t) + M}{(1+M)\tilde{s}_0^{out}(t)}\right] + \frac{2M(M-m)}{\tilde{s}_0^{out}(t) + M} - M\right\}. \tag{3.24}$$

Indeed, if we were to neglect the *secular* terms (*i.e.*, the terms proportional to τ in (3.21)), these solutions would also correctly satisfy the matching conditions, since one has:

$$\tilde{s}_1^{out}(0) = -\frac{1+m}{(1+M)^2}; \qquad \tilde{c}_1^{out}(0) = -\frac{M(1+2m-M)}{(1+M)^4}, \tag{3.25}$$

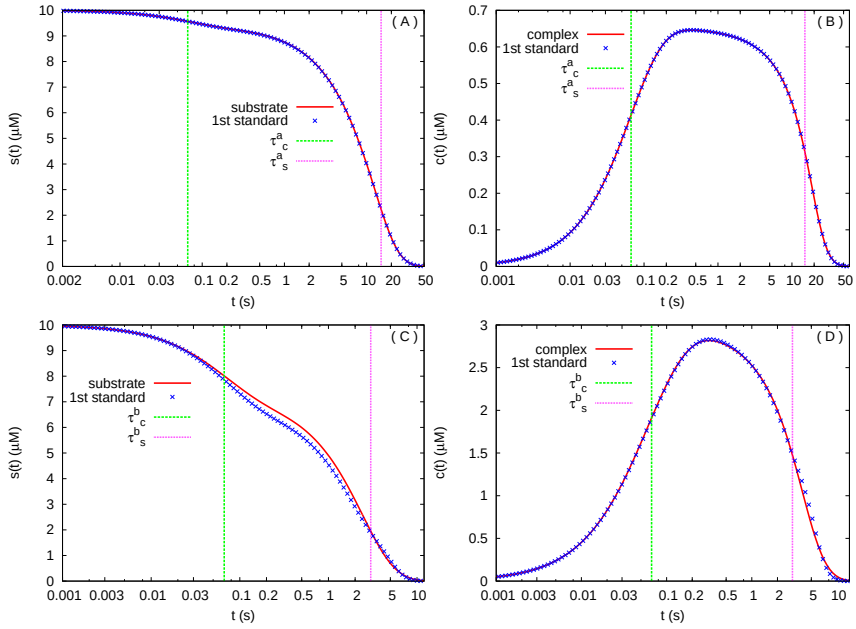

Figure 3.4: **Comparison between the numerical solutions and the 1st order perturbation expansion uniform approximations, for the two considered sets of ICs.** In A) and C) we present the behaviour of the concentrations of the substrate $s(t)$, whereas in B) and D) we present the one of the concentrations of the complex $c(t)$ for the a and b sets of ICs given by (3.8), respectively. Hence, in A) and B) we are in the case with $\varepsilon = \varepsilon^a = 0.1$, whereas in C) and D) we are in the one with $\varepsilon = \varepsilon^b = 0.5$. We plot both the numerical solutions of Eqs. (3.6), already shown in the previous figures, and the analytical solutions computed from the 1st order perturbation expansion uniform approximations (by using a standard numerical approximation to the Lambert function), as given by (3.28). We finally plot our corresponding rough evaluations of the two different time scales involved, too, with τ_s describing the substrate decay time and τ_c the complex saturation time. Notice that the time is in logarithmic scale.

that are the constant terms in the 1st order inner solutions.

From the point of view of the present work, it appears important to underline that one has, instead, to reasonably justify [2] the disappearance of the 1st order secular terms with the imposition of two term matching conditions, that involve the 1st order derivatives of the 0th order outer solutions, too. Moreover [2], one iteratively expects that, at higher orders in the expansion in ε, the divergences (*i.e.*, the expected presence of secular terms proportional to higher powers of τ in the inner solutions) could be absorbed by a possibly increasing number of terms in the corresponding matching conditions.

In detail, from Eqs. (3.11) one has:

$$\dot{\tilde{s}}_0^{out}(0) = -\frac{M-m}{(1+M)}; \qquad \dot{\tilde{c}}_0^{out}(0) = -\frac{M(M-m)}{(1+M)^3}; \tag{3.26}$$

Correspondingly, they turn out to be verified the two term matching conditions, reported in detail in [2], that are more or less resumable as:

$$\begin{cases} \lim_{\tau\to\infty} \left\{ \left[\tilde{s}_0^{in}(\tau) + \varepsilon \tilde{s}_1^{in}(\tau)\right] - \left[\tilde{s}_0^{out}(0) + t\dot{\tilde{s}}_0^{out}(0) + \varepsilon \tilde{s}_1^{out}(0)\right] \right\} &= 0 \\ \lim_{\tau\to\infty} \left\{ \left[\tilde{c}_0^{in}(\tau) + \varepsilon \tilde{c}_1^{in}(\tau)\right] - \left[\tilde{c}_0^{out}(0) + t\dot{\tilde{c}}_0^{out}(0) + \varepsilon \tilde{c}_1^{out}(0)\right] \right\} &= 0. \end{cases} \tag{3.27}$$

Noticeably, it was also proved [2] that these 1st order matching conditions apply on a time interval (the matching region) that ranges from ε to $\sqrt{\varepsilon}$.

Therefore, one takes as perturbation expansion uniform approximations to the solutions at the

1st order in ε:

$$\begin{cases} \tilde{s}_1^u(t) &= \left[\tilde{s}_0^{in}(t/\varepsilon) + \tilde{s}_0^{out}(t)\right] + \varepsilon\left[\tilde{s}_1^{in}(t/\varepsilon) + \tilde{s}_1^{out}(t)\right] + \\ &- \left[1 - \varepsilon\frac{1+m}{(1+M)^2} - \frac{M-m}{(1+M)}t\right] + O(\varepsilon^2) \\ \tilde{c}_1^u(t) &= \left[\tilde{c}_0^{in}(t/\varepsilon) + \tilde{c}_0^{out}(t)\right] + \varepsilon\left[\tilde{c}_1^{in}(t/\varepsilon) + \tilde{c}_1^{out}(t)\right] + \\ &- \left[\frac{1}{1+M} - \varepsilon\frac{M(1+2m-M)}{(1+M)^4} - \frac{M(M-m)}{(1+M)^3}t\right] + O(\varepsilon^2). \end{cases} \tag{3.28}$$

Let us remark that in [2], at pp. 315/316, it was proved that the error between these approximations and the correct solutions is smaller than $a\varepsilon^2$, by assuming a small enough ε, for $t \in [0, \infty)$, where a is a constant independent of both ε and t (*i.e.* the error is uniformly $O(\varepsilon^2)$ for $t \in [0, \infty)$). The statement is mainly based on the results obtained in [95, 31, 96].

We plot in [Fig. 3.4] our results on the substrate and complex concentrations, for the two considered sets of ICs, in comparison with the numerical solutions of the original problem, given by Eqs. (3.6) (the same curves as in the previous figures). The figures make evident that in the case a ([Fig. 3.4A] and [Fig. 3.4B]), *i.e.*, for the relatively small $\varepsilon = \varepsilon^a = 0.1$, the 1st order perturbation expansion uniform approximations are indistinguishable, within the plot precision (and within our numerical precision, too), from the correct solution on the whole relevant time interval, despite the particularly unfavourable situation, that is characterized both by the presence of three inflection points in the curve of the substrate and by a complex initial value, within the sQSSA, that is higher than its correct maximum.

On the other hand, in case b, *i.e.*, for $\varepsilon = \varepsilon^b = 0.5$, that is instead a quite large value, there is a naked-eye detectable difference between the approximation and the correct solution in the case of the substrate ([Fig. 3.4C]), for $t \sim 0.1 - 2\ s$. This region actually encompasses the matching one, that is expected to extend here from $t = \varepsilon^b/\delta^b = 0.1s$ to $t = \sqrt{\varepsilon^b}/\delta^b \simeq 0.14s$. Moreover, again in case b, when looking carefully at [Fig. 3.4D], one can notice that the approximation to the correct complex behaviour tends to zero still slightly too rapidly. In fact, also the maximum reached by the complex during its evolution is still slightly over-evaluated by the approximation. Both of these last observations, that concern the 1st order uniform approximation to the complex solution, within the perturbation expansion framework, will be made more evident by the following [Fig. 3.8C] and [Fig. 3.8D].

3.4 The SPDERG approach

It is worth noticing that, in the original works [26, 27], it was proposed, as a successful alternative to standard singular and reductive Perturbation Methods, a Renormalization Group approach to Singularly Perturbed Differential Equation in a more general context. Here, we limit ourselves to recall in some detail the discussion in [26, 27] in the case of boundary layer problems, that is the approach to which we refer by shortening with SPDERG, too.

As MM kinetics [29, 1, 2, 41], these problems [10, 27] are generally characterized by a boundary layer of a given small thickness (that in the case of MM kinetics is $O(\varepsilon)$, but that could also be, for instance, $O(\sqrt{\varepsilon})$), in which the solution is rapidly varying. Correspondingly, in order to predict the system's dynamics, one needs to solve singularly perturbed ODEs and one usually resorts to the standard perturbation expansion method.

Let us attempt to sketch the general SPDERG procedure in these cases, by following in particular the examples of its application to boundary layer problems given in [27], one of which we will recall in detail, in order to better illustrate the method:

- One faces the singular 2nd order ODE to be obeyed by the outer solution $y(t)$;

- One makes the transformation $t \to \tau = t/\varepsilon$, thereby solving the 2nd order ODE to be obeyed by the inner solution $Y(\tau)$, with given ICs (*i.e.*, Cauchy data) at $\tau = \tau_0$, and obtaining this last function;

- One identifies the part to be renormalized of the function $Y(\tau)$, *i.e.*, $Y_{div}(\tau)$. One moreover identifies, within $Y_{div}(\tau)$, the appropriate *bare* integration constants, that have to contain the ICs;

- When appropriately choosing these constants, their replacement with the renormalized ones, that appear for the moment to be λ-depending, together with the replacement $(\tau - \tau_0) \to (\tau - \lambda) + (\lambda - \tau_0)$, allows to get the correct $Y_{div}(\tau, \lambda)$. Indeed, this function depends on τ in the same way as the original $Y_{div}(\tau)$. The differences are: i) the replacement $\tau_0 \to \lambda$; ii) the λ-dependence of the integration constants;

- One renormalizes $Y_{div}(\tau, \lambda)$ by imposing to this function to be independent of the *arbitrary* time λ, *i.e.*, $dY_{div}(\tau, \lambda)/d\lambda = 0$. This last requirement allows to get the ODEs to be obeyed by the *renormalized* integration constants.

- One makes the final transformation $\lambda \to \tau$;

- One adds to the obtained renormalized function the remaining part of the original inner solution $Y(\tau)$, in which the integration constants are not renormalized, then one imposes the original ICs and finally one looks at the result as a function of $\tau = t/\varepsilon$.

By this way, one ends up with a solution that indeed, usually:

- Contains the outer contribution up to the leading order in ε, too;

- Does not contain secular terms, hence does not diverge for $t \to \infty$;

- Does not require to cancel terms that otherwise would appear twice;

- Can be, for these reasons, regarded as the SPDERG uniform approximation to the correct solution;

- Can be expected to approximate better the correct solution to the original problem than the perturbation expansion uniform approximation at the leading order, at least in the matching region, since it contains the inner contribution at the following one.

In order to make a remark about similarities and differences with the other standard RG techniques that we roughly recalled in Section 3.1, as applied in particular in Statistical Field Theory, we notice that here the integration constants are the equivalent of the key parameters of the theory, whereas λ plays the role of a (large time) cutoff. From this point of view, the ODEs to be obeyed by the integration constants are the equivalent of the RG scaling equations. Indeed, by solving them, one is contemporaneously eliminating λ, therefore the divergences. Within this analogy, the system is already at its fixed point, since at the end of the recalled appropriate equivalent of the coarse graining / rescaling procedure, one finds a function $Y_{div}(\tau, \lambda)$ with the same dependence on τ as $Y_{div}(\tau)$. In particular, the implied scale invariance appears to be interpreted as the invariance with respect to λ.

To recall more in detail the discussion in [27], one chooses the part to be renormalized, $Y_{div}(\tau)$, of the original inner solution, $Y(\tau)$, by requiring that it contain the ICs, the other leading order terms and the terms that do not tend to zero or to a constant value for $(\tau - \tau_0) \to \infty$ (*i.e.*, the secular terms). Moreover, the bare integration constants, let us say (following [27] also in the notation, as previously in this Section) $A_0(\tau_0)$ and $B_0(\tau_0)$, are renormalized as: $A_0(\tau_0) \to Z_1 A(\lambda)$, $B_0(\tau_0) \to Z_2 B(\lambda)$. In

fact, the renormalization constants, Z_1 and Z_2, depend both on τ_0 and λ, their basic role being to change, in the part to be renormalized of the function, the initial dependence on τ_0 of the ICs in a dependence on the arbitrary time λ.

In particular, they are assumed to have the expansion (with $a_0 = b_0 = 1$):

$$Z_1 = \sum_{n=0}^{\infty} a_n(\tau_0, \lambda)\varepsilon^n; \qquad Z_2 = \sum_{n=0}^{\infty} b_n(\tau_0, \lambda)\varepsilon^n. \tag{3.29}$$

Since one can always take $(\tau - \tau_0) = (\tau - \lambda) + (\lambda - \tau_0)$, the secular terms in $(\lambda - \tau_0)$ are correspondingly absorbable in appropriate redefinitions of A_0 and B_0, that have to be made by correctly choosing the coefficients $\{a_n\}$ and $\{b_n\}$. As it will become clear in the application to MM kinetics that we are going to present, one can usually safely expect to be able to absorb in the coefficients of the renormalization constants also other possible secular terms, for instance those corresponding to higher powers of $(\tau - \tau_0)$. In particular, we will see in the present 2nd order calculations that, when writing $(\tau - \tau_0)^2 = (\tau - \lambda)^2 + (\lambda - \tau_0)^2 + 2(\tau - \lambda)(\lambda - \tau_0)$, only the second term is to be absorbed, whereas the last one is cancelled by the previously chosen 1st order renormalization coefficient, in the same way as in a case that is considered in [27] (see Section B, in that work).

Therefore [26, 27], the basic hypothesis is that the bare quantities need to be renormalized in such a way to get a solution independent of the arbitrary time λ, as it has reasonably to be. This is shown explicitly in the example that we are going to give, taken from [27]. Evidently, λ plays a key role in the whole approach and one can correspondingly think to it also as the equivalent of the *unknown time* at which the matching needs to be obeyed. From this point of view, we notice, therefore, an analogy with the imposition of the matching conditions, in the standard perturbation expansion method. This analogy can perhaps, first of all, be useful to the readers without particular knowledge of the Renormalization Group techniques. In fact, this analogy can be even more advantageous in order to better understand similarities and differences between the SPDERG approach and the perturbation expansion method. Nevertheless, it is not to be taken too rigorously, at least in cases as the present one, where, already at the 1st order, the matching requires two term conditions and one expects that the correct matching could require even more terms at higher orders [2].

Let us give a detailed example, taken from [27] (see Section C in that work), of an application of the SPDERG approach to a nonlinear boundary layer problem, as in the present case of MM kinetics.

We consider the 2nd order ODE:

$$\begin{aligned} \varepsilon\ddot{y}(t) + 2\dot{y}(t) + e^{y(t)} &= 0 \\ y(0) = y(1) &= 0. \end{aligned} \tag{3.30}$$

This example is characterized by a boundary layer, that is in particular of thickness ε (once again, as in the present case of MM kinetics), at $t = 0$. In order to study the transient phase, we rescale the independent variable: $\tau = t/\varepsilon$ and rename the dependent variable: $Y(\tau) = y(t)$. Thus we have:

$$\frac{d^2Y(\tau)}{d\tau^2} + 2\frac{dY(\tau)}{d\tau} = -\varepsilon e^{Y(\tau)} \ . \tag{3.31}$$

This is the analogous of the inner equation. We solve the equation by means of the expansion $Y(\tau) = Y_0 + \varepsilon Y_1(\tau) + O(\varepsilon^2)$ and we obtain, in the large τ limit:

$$Y(\tau) \sim A_0 + B_0 e^{-2\tau} - \varepsilon \left[\frac{1}{2}e^{A_0}(\tau - \tau_0) + \mathcal{R}\right] + O(\varepsilon^2) \ , \tag{3.32}$$

where A_0 and B_0 are integration constants and $\mathcal{R}$ contains all the terms asymptotically regular for $(\tau - \tau_0) \to +\infty$. In order to outline the difference with the application of the SPDERG method in the present work, it is interesting to notice that here one faces directly a second order differential equation,

that is solved by introducing two appropriate integration constants. Hence, one has consistently a single time τ_0 at which these two integration constants need to satisfy the two given ICs, at the end of the procedure.

In detail, the part to be renormalized of the function, $Y_{div}(\tau)$, is the one in which the contribution due to $\mathcal{R}$ is neglected. Its divergence can be controlled by renormalizing the two integration constants as $A_0(\tau_0) \to Z_1 A(\lambda)$ and $B_0(\tau_0) \to Z_2 B(\lambda)$, respectively. The renormalization constants, Z_1 and Z_2, are assumed to depend both on τ_0 and λ, with the aim of changing, in $Y_{div}(\tau)$, the initial dependence on τ_0 of the ICs in a dependence on the arbitrary time λ.

Generally, for a boundary layer of thickness ε, one moreover chooses, for Z_1 and Z_2, the expansion (with $a_0 = b_0 = 1$) that we already reported in (3.29). Hence, at the first order in ε:

$$Z_1 = 1 + a_1(\tau_0, \lambda)\varepsilon + O(\varepsilon^2); \qquad Z_2 = 1 + b_1(\tau_0, \lambda)\varepsilon + O(\varepsilon^2). \tag{3.33}$$

In this example, it is important to outline that the term proportional to $e^{-\tau}$ is anyway suppressed in the large τ limit. From this point of view, the case is similar to the one that we will encounter in the context of our SPDERG study of the complex. Therefore, at the next step of the procedure, *i.e.*, when replacing $(\tau - \tau_0)$ with $(\tau - \lambda) + (\lambda - \tau_0)$, the only constant to be fixed is a_1. One chooses its value exactly in order to obtain a τ_0-independent function, which implies:

$$a_1 = \frac{1}{2} e^{A_0} (\lambda - \tau_0)/A_0. \tag{3.34}$$

Consistently, one arrives at the renormalized perturbation result

$$Y_{div}(\tau, \lambda) = A(\lambda) + B(\lambda) e^{-2\tau} - \varepsilon \frac{1}{2} e^{A(\lambda)} (\tau - \lambda) + O(\varepsilon^2) \ . \tag{3.35}$$

Imposing the renormalization group condition of invariance $dY/d\lambda = 0$ brings to the equations, up to order ε:

$$\begin{aligned} \frac{dA(\lambda)}{d\lambda} + \varepsilon \frac{1}{2} e^{A(\lambda)} &= 0 \\ \frac{dB(\lambda)}{d\lambda} &= 0 \ . \end{aligned} \tag{3.36}$$

Therefore, one obtains:

$$A(\lambda) = \log\left(\frac{2}{\varepsilon\lambda + C_1}\right) \ ; \qquad B(\lambda) = C_2 \ , \tag{3.37}$$

where C_1 and C_2 are two integration constants to be determined by means of the boundary values. To better outline the difference with our SPDERG results on the complex, we anticipate that, in our context, a constant renormalized complex IC will imply instead a zero value for this quantity, hence the possibility to neglect the terms proportional to it.

Equating λ and τ in (3.35) and rewriting $t = \varepsilon\tau$, we obtain the SPDERG uniform expansion:

$$y(t) = \log\left(\frac{2}{t + C_1}\right) + C_2 e^{-2t/\varepsilon} + O(\varepsilon) \ . \tag{3.38}$$

Finally, by imposing the boundary conditions given in (3.30) and by taking into account the limit $\varepsilon \to 0^+$, we arrive at the approximation:

$$y(t) = \log\left(\frac{2}{t+1}\right) + \log(2) e^{-2t/\varepsilon} + O(\varepsilon) \ . \tag{3.39}$$

The result reproduces, at the leading order, the solution obtained in [10], (Example 1, pp. 463–464, formula (9.7.6)) in terms of singular perturbations and matching techniques.

3.5 SPDERG results beyond the sQSSA

3.5.1 First order contribution

Here we use, for the first time to our knowledge, the SPDERG approach to study MM kinetics, beyond the sQSSA, whose known uniform approximations we already recalled in Section 3.3.1. In fact, we introduce an ad hoc new way of applying the SPDERG to the present case, that is different from the one considered in particular in [27].

In order to explain why our different procedure seems useful here, we start by noticing that one can rewrite the system, given by (3.15), for the inner solutions, in the form of a single 2nd order ODE (see [83, 84, 97] for the first studies of MM kinetics in terms of a 2nd order ODE), to be obeyed by the dimensionless substrate concentration:

$$\ddot{\tilde{s}}^{in}(\tau)-\frac{\left[\dot{\tilde{s}}^{in}(\tau)\right]^2}{\tilde{s}^{in}(\tau)+m}+\left[\tilde{s}^{in}(\tau)+M\right]\dot{\tilde{s}}^{in}(\tau)+\varepsilon\left[\frac{m}{\tilde{s}^{in}(\tau)+m}\dot{\tilde{s}}^{in}(\tau)+(M-m)\tilde{s}^{in}(\tau)\right]=0, \tag{3.40}$$

with ICs (Cauchy data) $\tilde{s}^{in}(0)=1$ and $\dot{\tilde{s}}^{in}(0)=-\varepsilon$. Analogously, one can write the same system, given by (3.15), in the form of a single 2nd order ODE, to be obeyed by the dimensionless complex concentration:

$$\left[1-\tilde{c}^{in}(\tau)\right]\ddot{\tilde{c}}^{in}(\tau)+\left[\dot{\tilde{c}}^{in}(\tau)\right]^2+M\tilde{c}^{in}(\tau)\dot{\tilde{c}}^{in}(\tau)+\varepsilon\left[1-\tilde{c}^{in}(\tau)\right]^2\left[\dot{\tilde{c}}^{in}(\tau)+(M-m)\tilde{c}^{in}(\tau)\right]=0, \tag{3.41}$$

with ICs (Cauchy data) $\tilde{c}^{in}(0)=0$ and $\dot{\tilde{c}}^{in}(0)=1$.

Let us, in the meantime, observe that, at the 1st order in ε, the equations are satisfied by $\tilde{s}^{in}(\tau)=\tilde{s}_0^{in}(\tau)+\varepsilon\tilde{s}_1^{in}(\tau)$ and $\tilde{c}^{in}(\tau)=\tilde{c}_0^{in}(\tau)+\varepsilon\tilde{c}_1^{in}(\tau)$, with $\tilde{s}_0^{in}(\tau)$, $\tilde{c}_0^{in}(\tau)$ given by (3.17) and $\tilde{s}_1^{in}(\tau)$, $\tilde{c}_1^{in}(\tau)$ given by (3.21), respectively. On general bases, one expects that the solutions of these second order ODEs are the same as the ones of the system given by Eqs. (3.15), at each order in ε.

Nevertheless, the main difficulty, when attempting to apply the SPDERG approach scheme of [27] to the 2nd order ODE for the substrate, Eq. (3.40), or, equivalently, to the 2nd order one for the complex, Eq. (3.41), is that one faces the problem already encountered in [41], *i.e.*, that this scheme leaves with some freedom in the choice of the integration constants to be renormalized. In fact, to identify the most appropriate ones is seemingly more complicated in the case of MM kinetics than in the other boundary layer problems studied in [27].

For this reason we start by studying the SPDERG dimensionless substrate and complex concentrations, $\tilde{s}^{rg}(\tau)$ and $\tilde{c}^{rg}(\tau)$, that are solutions of the system given by Eqs. (3.15). The difference, with respect to the $\tilde{s}^{in}(\tau)$ and $\tilde{c}^{in}(\tau)$ that we previously recalled, within the perturbation expansion method, is that, here, the ICs are given at a time τ_0, that we assume to be in principle different from zero, as $\tilde{s}^{rg}(\tau_0)=\tilde{s}^*$ and $\tilde{c}^{rg}(\tau_0)=\tilde{c}^*$, respectively. Correspondingly, we moreover take these ICs, hence $\tilde{s}^*$ and $\tilde{c}^*$, as the bare quantities to be renormalized, whereas the original ICs of the problem, *i.e.*, $\tilde{s}(0)=1$ and $\tilde{c}(0)=0$, will be taken into account after the renormalization procedure.

For the sake of clarity, at the cost of being somehow repetitive, at the 1st order in ε, we look again for solutions in the form:

$$\begin{cases} \tilde{s}^{rg}(\tau)=\tilde{s}_0^{rg}(\tau)+\varepsilon\tilde{s}_1^{rg}(\tau) \\ \tilde{c}^{rg}(\tau)=\tilde{c}_0^{rg}(\tau)+\varepsilon\tilde{c}_1^{rg}(\tau). \end{cases} \tag{3.42}$$

In these formulas, $\tilde{s}_0^{rg}(\tau)$ and $\tilde{c}_0^{rg}(\tau)$ solve as usual the 0th order system for the inner solutions, but the ICs are given at $\tau=\tau_0$ and their values are $\tilde{s}_0^{rg}(\tau_0)=\tilde{s}_0^*$ and $\tilde{c}_0^{rg}(\tau_0)=\tilde{c}_0^*$. These solutions are:

$$\begin{cases} \tilde{s}_0^{rg}(\tau) &= \tilde{s}_0^* \\ \tilde{c}_0^{rg}(\tau) &= \tilde{c}_0^* e^{-(\tilde{s}_0^*+M)(\tau-\tau_0)}+\dfrac{\tilde{s}_0^*}{\tilde{s}_0^*+M}\left[1-e^{-(\tilde{s}_0^*+M)(\tau-\tau_0)}\right]. \end{cases} \tag{3.43}$$

On the other hand, $\tilde{s}_1^{rg}(\tau)$ and $\tilde{c}_1^{rg}(\tau)$ need to solve the 1st order system (that is slightly more complicated than Eqs. (3.20) because of the presence of $\tilde{s}_0^* \neq 1$):

$$\left\{ \begin{array}{rcl} \dot{\tilde{s}}_1^{rg}(\tau) & = & (\tilde{s}_0^* + m)\tilde{c}_0^{rg}(\tau) - \tilde{s}_0^* \\ \dot{\tilde{c}}_1^{rg}(\tau) & = & -(\tilde{s}_0^* + M)\tilde{c}_1^{rg}(\tau) - [\tilde{c}_0^{rg}(\tau) - 1]\, \tilde{s}_1^{rg}(\tau), \end{array} \right. \tag{3.44}$$

with ICs, at $\tau = \tau_0$, $\tilde{s}_1^{rg}(\tau_0) = \tilde{s}_1^*$ and $\tilde{c}_1^{rg}(\tau_0) = \tilde{c}_1^*$. We obtain, in the case of the substrate:

$$\begin{array}{rcl} \tilde{s}_1^{rg}(\tau) & = & \tilde{s}_1^* - (M-m)\dfrac{\tilde{s}_0^*}{\tilde{s}_0^* + M}(\tau - \tau_0) + \\ & - & \dfrac{\tilde{s}_0^* + m}{(\tilde{s}_0^* + M)^2}\left[\tilde{s}_0^* - \tilde{c}_0^*(\tilde{s}_0^* + M)\right]\left[1 - e^{-(\tilde{s}_0^* + M)(\tau - \tau_0)}\right], \end{array} \tag{3.45}$$

whereas the solution for the complex, $\tilde{c}_1^{rg}(\tau)$, is given in Appendix B.1.

Let us continue by studying $\tilde{s}^{rg}(\tau)$. Following [27], we look at $\tilde{s}_{div}^{rg}(\tau)$, that contains only the substrate IC (*i.e.*, the terms that give $\tilde{s}^{rg}(\tau_0) = \tilde{s}_0^* + \varepsilon\tilde{s}_1^*$), the other possible leading terms at this order (that are absent, in the present case) and the secular terms (here, the one proportional to $(\tau - \tau_0)$ in $\tilde{s}_1^{rg}(\tau)$). Therefore, we write:

$$\tilde{s}^{rg}(\tau) = \tilde{s}_{div}^{rg}(\tau) + \varepsilon\mathcal{R}_1^s(\tau) + O(\varepsilon^2), \tag{3.46}$$

by grouping in $\mathcal{R}_1^s$ all the sub-leading terms that tend to zero or to a constant for $(\tau - \tau_0) \to \infty$ and that do not need to be renormalized. One has:

$$\tilde{s}_{div}^{rg}(\tau) = \tilde{s}_0^* + \varepsilon\tilde{s}_1^* - \varepsilon(M-m)\frac{\tilde{s}_0^*}{\tilde{s}_0^* + M}(\tau - \tau_0). \tag{3.47}$$

As previously recalled, the SPDERG approach [26, 27], at this point, is based on writing $(\tau - \tau_0) = (\tau - \lambda) + (\lambda - \tau_0)$, by correspondingly assuming, in the present case, that the secular term proportional to $(\lambda - \tau_0)$ can be absorbed in an appropriate redefinition of the bare substrate IC. In detail, by labeling $\tilde{s}_0^{rg*}$ and $\tilde{s}_1^{rg*}$, the contributions to the renormalized IC at the 0th and at the 1st order, respectively, we put $\tilde{s}_0^* = Z_{s_0}\tilde{s}_0^{rg*}(\lambda)$ and $\tilde{s}_1^* = Z_{s_1}\tilde{s}_1^{rg*}(\lambda)$. The renormalization constants, Z_{s_0} and Z_{s_1}, are assumed to have the same ε-expansions as the ones given by (3.29). These expansions, at the 1st order in ε, imply $\tilde{s}_0^* = [1 + \varepsilon z_{s_0,1}(\tau_0, \lambda)]\tilde{s}_0^{rg*}(\lambda)$ and $\tilde{s}_1^* = \tilde{s}_1^{rg*}(\lambda)$. Thus, the present secular term can be absorbed by choosing the first free coefficient in the expansions as:

$$z_{s_0,1}(\tau_0, \lambda) = \frac{(M-m)}{\tilde{s}_0^{rg*} + M}(\lambda - \tau_0). \tag{3.48}$$

Correspondingly, we end up to study:

$$\tilde{s}_{div}^{rg}(\tau, \lambda) = \tilde{s}_0^{rg*}(\lambda) + \varepsilon\tilde{s}_1^{rg*}(\lambda) - \varepsilon(M-m)\frac{\tilde{s}_0^{rg*}(\lambda)}{\tilde{s}_0^{rg*}(\lambda) + M}(\tau - \lambda). \tag{3.49}$$

Hence, by imposing the scaling condition [26, 27], $d\tilde{s}_{div}^{rg}(\tau, \lambda)/d\lambda = 0$, we get the 1st order ODEs, to be obeyed by the two terms that contribute to the renormalized IC.

In detail, at the 1st order in ε, *i.e.*, by disregarding in particular the part of the derivative of the last term that contains $d\tilde{s}_0^{rg*}(\lambda)/d\lambda$ (it will contribute to the 2nd order, as we show in the following Section 3.5.2), we obtain:

$$\frac{d\tilde{s}_0^{rg*}(\lambda)}{d\lambda} + \varepsilon\frac{d\tilde{s}_1^{rg*}(\lambda)}{d\lambda} + \varepsilon(M-m)\frac{\tilde{s}_0^{rg*}(\lambda)}{\tilde{s}_0^{rg*}(\lambda) + M} = 0. \tag{3.50}$$

Then, we separate the contributions by correctly taking into account, in advance, that, at the next step, one makes the transformation $\lambda \to \tau = t/\varepsilon$, and we get:

$$\begin{cases} \dfrac{d\tilde{s}_0^{rg*}(\lambda)}{d\lambda} &= -\varepsilon(M-m)\dfrac{\tilde{s}_0^{rg*}(\lambda)}{\tilde{s}_0^{rg*}(\lambda)+M} \\ \dfrac{d\tilde{s}_1^{rg*}(\lambda)}{d\lambda} &= 0. \end{cases} \tag{3.51}$$

Consistently, when indeed one makes the further transformation [27] $\lambda \to \tau = t/\varepsilon$, the first of these equations is just the ODE to be obeyed by the 0th order outer dimensionless substrate concentration, that we encountered in Eqs. (3.11). Its solution is therefore $\tilde{s}_0^{rg*}(t) = \tilde{s}_0^{out}(t)$, that is given by (3.12), by means of the Lambert function, and that satisfies the IC $\tilde{s}_0^{rg*}(0) = 1$, too.

Clearly, this result already suggests that, also in the case of MM kinetics, the present ad hoc new way of applying the SPDERG approach turns out to be able to reproduce the leading order terms of the perturbation expansion uniform approximations.

On the other hand, the result $d\tilde{s}_1^{rg*}(\lambda)/d\lambda = 0$ is, here, to be interpreted as the verification that $\tilde{s}_1^{rg*}$ can be neglected at this order. In fact, we find $\tilde{s}_1^{rg*} = const = 0$ by imposing the original IC. More correctly, the bare IC, that is fixed, here, to the value $\tilde{s}(0) = 1$, should be imposed on the solution at the end, but this is not influential in the present case, since one generally has $\mathcal{R}(0) = 0$, for the appropriately calculated contribution of the parts not to be renormalized of the inner solutions.

Finally, we get the renormalized result:

$$\tilde{s}_1^{rg,u}(t) = \tilde{s}_0^{out}(t) + \varepsilon\mathcal{R}_1^s(t/\varepsilon) + O(\varepsilon). \tag{3.52}$$

Notice that, when we were considering $\tilde{s}^{rg}(\tau)$ as the inner solution, in (3.46), before applying the SPDERG approach, we knew that it was correct up to order $O(\varepsilon^2)$. Here, instead, the renormalized $\tilde{s}_1^{rg,u}(t)$ (that for this reason we label *rg,u*, by moreover making explicit that it is a 1st order approximation) is to be interpreted as the SPDERG uniform approximation to the original problem. Correspondingly, one has to bear in mind that, from this point of view, it is correct only up to order $O(\varepsilon)$. Actually, the renormalized $\tilde{s}_1^{rg,u}(t)$ in (3.52) is only expected [26, 27] to contain the leading order terms of the outer solution. It is indeed the case, by its comparison with (3.28).

Furthermore, the terms in $\mathcal{R}_1^s$ (*i.e.*, the ones that do not need to be renormalized) are to be evaluated in $\tau_0 = 0$, by using the correct ICs, namely, the bare ones (*i.e.*, $\tilde{s}_0^* = 1$, $\tilde{s}_1^* = 0$ and $\tilde{c}_0^* = 0$). Thus, as expected [27], the SPDERG 1st order uniform approximation to the solution for the dimensionless substrate concentration does instead contain the 1st order terms of the perturbation expansion uniform approximation, *i.e.*, the ones that originate from $\tilde{s}_1^{in}(t/\varepsilon)$. On the other hand, the usual asymptotic behaviour of the solution, that one finds when applying the recalled perturbation expansion method to MM kinetics (*i.e.*, $\lim_{t\to\infty}\tilde{s}(t) = 0$), is not verified here (one has $\lim_{t\to\infty}\tilde{s}_1^{rg,u}(t) = O(\varepsilon)$). We better discuss this failure of the present application of the SPDERG approach in the following Section 3.5.3, by proposing a way to overcome it, too.

Let us now study the complex. First of all, despite the presence of a large number of terms in (B.1) (the formula is given in Appendix B.1, with the coefficients reported in (B.2)), that describes the behaviour of the 1st order inner dimensionless complex concentration, within the SPDERG approach, $\tilde{c}_1^{rg}(\tau)$, the part of the function to be renormalized up to the 1st order in ε is as manageable as in the case of the substrate.

In detail, we write the same kind of expression as in (3.46):

$$\tilde{c}^{rg}(\tau) = \tilde{c}_{div}^{rg}(\tau) + \varepsilon\mathcal{R}_1^c(\tau) + O(\varepsilon^2), \tag{3.53}$$

by collecting in $\mathcal{R}_1^c$ all the sub-leading terms that tend to zero or to a constant for $(\tau - \tau_0) \to \infty$. Correspondingly, we find:

$$\tilde{c}_{div}^{rg}(\tau) = (\tilde{c}_0^* + \varepsilon\tilde{c}_1^*)e^{-(\tilde{s}_0^*+M)(\tau-\tau_0)} + \frac{\tilde{s}_0^*}{\tilde{s}_0^*+M} - \varepsilon\frac{M(M-m)}{(\tilde{s}_0^*+M)^3}\tilde{s}_0^*(\tau-\tau_0). \tag{3.54}$$

In fact, here, we considered explicitly also the terms that give the original complex IC, (*i.e.*, $\tilde{c}^{rg}_{div}(0) = \tilde{c}^*_0 + \varepsilon\tilde{c}^*_1$). Nevertheless, both the 0th order contribution and the 1st order one, to this IC, are exponentially suppressed for $(\tau - \tau_0) \to \infty$. In particular from this point of view, the present case is similar to the example of the nonlinear boundary layer problem given in [27], that we previously recalled. Correspondingly, $\tilde{c}^{rg*}_0(\lambda)$, when renormalizing, has to obey the ODE $d\tilde{c}^{rg*}_0(\lambda)/d\lambda = 0$. Hence, one finds $\tilde{c}^{rg*}_0(\lambda) = const = 0$ when imposing the original (bare) IC (fixed, here, to $\tilde{c}(0) = 0$), at the end of the renormalization procedure. Namely, the contribution proportional to this term can be neglected. For the same reason, one also expects $\tilde{c}^{rg*}_1(\lambda) = const = 0$, a result that is even more predictable at this order, since we already verified that $\tilde{s}^{rg*}_1$ gives no contribution in the case of the substrate. Correspondingly, we assume that the contribution proportional to this last term can be neglected, too.

Accordingly, the function to be studied is further reduced to:

$$\tilde{c}^{rg}_{div}(\tau)\Big|_{\substack{\tilde{c}^*_0=0\\ \tilde{c}^*_1=0}} = \frac{\tilde{s}^*_0}{\tilde{s}^*_0 + M} - \varepsilon(M-m)\frac{M\tilde{s}^*_0}{(\tilde{s}^*_0 + M)^3}(\tau - \tau_0). \tag{3.55}$$

As already recalled, in the context of the standard perturbation expansion method, the matching condition that is verified by the dimensionless substrate concentration does consistently also satisfy the matching of the inner and outer dimensionless complex concentrations. Alternatively, one can also remind that the SPDERG approach is usually applied to a 2nd order ODE (*i.e.*, here, generally speaking, we could have chosen, equivalently, either the one for the substrate, Eq. (3.40), or the one for the complex, Eq. (3.41)).

For these reasons, on the basis of the previously sketched analogy, too, we find not surprising that, when making the transformations $\left\{\tau_0 \to \lambda, \tilde{s}^*_0 \to \tilde{s}^{rg*}_0(\lambda)\right\}$, the scaling condition $d\tilde{c}^{rg}_{div}(\tau,\lambda)/d\lambda = 0$ gives, once again, the same ODE to be obeyed by $\tilde{s}^{rg*}_0(\lambda)$ as the one that we just found in the study of $\tilde{s}^{rg}_{div}(\tau,\lambda)$ (*i.e.*, Eqs. (3.51)). Hence (with the transformation $\lambda \to \tau = t/\varepsilon$), we get, once again, $\tilde{s}^{rg*}_0(t) = \tilde{s}^{out}_0(t)$. In particular, one can check that, since $d[\tilde{s}^*_0/(\tilde{s}^*_0 + M)]/d\tilde{s}^*_0 = M/(\tilde{s}^*_0 + M)^2$, when renormalizing the bare 0th order contribution to the substrate IC, $\tilde{s}^*_0$, by using the value of the coefficient $z_{s_0,1}$, that is already fixed by (3.48), the contribution of the first term, in $\tilde{c}^{rg}_{div}(\tau,\lambda)$, to the order ε, is exactly equal to $\varepsilon(\lambda - \tau_0)M(M-m)\tilde{s}^{rg*}_0/(\tilde{s}^{rg*}_0 + M)^2$ and, therefore, it suitably absorbs the secular term that is present in this case.

By recalling that $\tilde{c}^{out}_0(t) = \tilde{s}^{out}_0(t)/(\tilde{s}^{out}_0(t) + M)$ from Eqs. (3.11), the obtained SPDERG 1st order uniform approximation for the complex is:

$$\tilde{c}^{rg,u}_1(t) = \tilde{c}^{out}_0(t) + \varepsilon\mathcal{R}^c_1(t/\varepsilon) + O(\varepsilon), \tag{3.56}$$

with, as for the substrate, the terms in $\mathcal{R}^c_1$ to be evaluated in $\tau_0 = 0$, by using the original ICs (*i.e.*, $\tilde{s}^*_0 = 1$ and $\tilde{s}^*_1 = \tilde{c}^*_0 = \tilde{c}^*_1 = 0$).

Writing explicitly the contribution of the different terms, we obtain the following SPDERG 1st order uniform approximations to the correct solutions for the time behaviour of the dimensionless substrate concentration and of the dimensionless complex concentration, in MM kinetics, beyond the

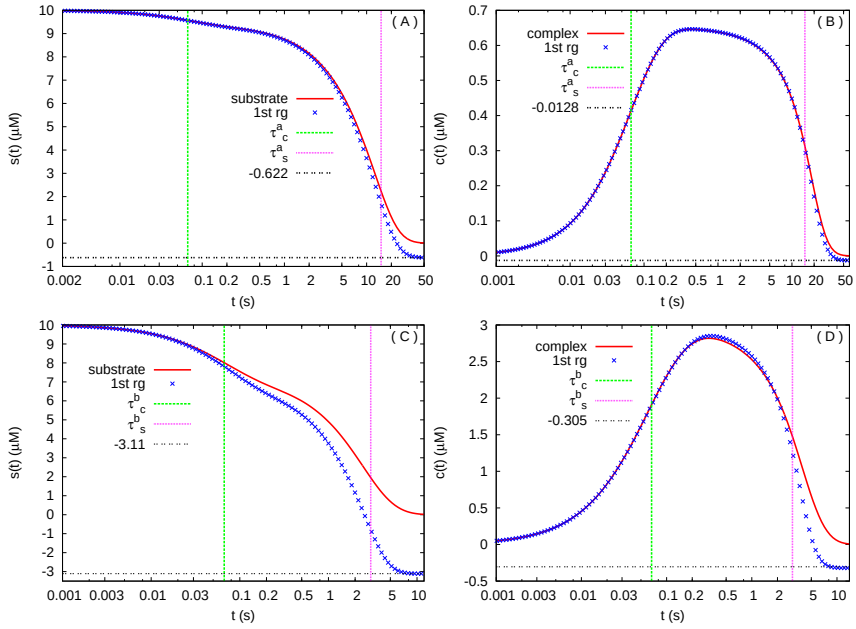

Figure 3.5: **Comparison between the numerical solutions and the 1st order SPDERG uniform approximations, for the two considered sets of ICs.** In A) and C) we present the behaviour of the concentrations of the substrate $s(t)$, whereas in B) and D) we present the one of the concentrations of the complex $c(t)$, for the a and b sets of ICs given by (3.8), respectively. Hence, in A) and B) we are in the case with $\varepsilon = \varepsilon^a = 0.1$, whereas in C) and D) we are in the one with $\varepsilon = \varepsilon^b = 0.5$. We plot both the numerical solutions of Eqs. (3.6), already shown in the previous figures, and the analytical solutions computed from the SPDERG 1st order uniform approximations (with a standard numerical approximation to the Lambert function), as given by (3.57). We plot moreover the (physically meaningless) limits for $t \to \infty$ of the analytical solutions: $s_0^a \tilde{s}_{1,\infty}^{rg,u} \simeq -0.622$ (μM), $e_0^a \tilde{c}_{1,\infty}^{rg,u} \simeq -0.0122$ (μM) and $s_0^b \tilde{s}_{1,\infty}^{rg,u} \simeq -3.11$ (μM), $e_0^b \tilde{c}_{1,\infty}^{rg,u} \simeq -0.305$ (μM), respectively. We finally plot our corresponding rough evaluations of the two different time scales involved, too, with τ_s describing the substrate decay time and τ_c the complex saturation time. Notice that the time is in logarithmic scale.

sQSSA, respectively (with ω the Lambert function):

$$
\begin{cases}
\tilde{s}_1^{rg,u}(t) &= M\omega(e^{-(M-m)t/M+1/M}/M) + \\
&- \varepsilon \dfrac{1+m}{(1+M)^2}\left[1-e^{-(1+M)t/\varepsilon}\right] + O(\varepsilon) \\
\tilde{c}_1^{rg,u}(t) &= \dfrac{\omega(e^{-(M-m)t/M+1/M}/M)}{\omega(e^{-(M-m)t/M+1/M}/M)+1} - \dfrac{1}{1+M} e^{-(1+M)t/\varepsilon} + \\
&- \varepsilon \dfrac{M(1+2m-M)}{(1+M)^4}\left[1-e^{-(1+M)t/\varepsilon}\right] + \\
&- \varepsilon \left[\dfrac{(1-M)(1+m)}{(1+M)^3}(t/\varepsilon) + \dfrac{M-m}{(1+M)^2}\dfrac{(t/\varepsilon)^2}{2}\right] e^{-(1+M)t/\varepsilon} + \\
&+ \varepsilon \dfrac{(1+m)}{(1+M)^4} e^{-(1+M)t/\varepsilon}\left[1-e^{-(1+M)t/\varepsilon}\right] + O(\varepsilon).
\end{cases}
\tag{3.57}
$$

Here, we stress once again that these SPDERG uniform approximations are correct only up to $O(\varepsilon)$, since they lack the first order contributions of the outer solutions.

We plot in [Fig. 3.5] our corresponding results for the two considered sets of ICs, as usual, in comparison with the numerical solutions of the original problem, given by Eqs. (3.6) (the same curves as in the previous figures). The figure shows that, as expected, these results approximate more correctly the numerical solutions than the perturbation expansion 0th order uniform approximations

([Fig. 3.2A], [Fig. 3.3A] and [Fig. 3.2C], [Fig. 3.3B] for the results for the a and b sets of ICs, respectively). Nonetheless, these approximations are not yet as much correct as the perturbation expansion 1st order ones ([Fig. 3.4]).

Moreover, the figure also makes evident that, differently from the perturbation expansion results, the SPDERG approximations fail in particular in the large time region. Actually, also in the examples given in [27], the SPDERG was shown to give usually better results than the standard method in the matching region. Nevertheless, here we face the problem that both the substrate and the complex approach (physically meaningless) $O(\varepsilon)$ values for $t \to \infty$. In fact, from (3.57), one has $\lim_{t\to\infty} \tilde{s}_1^{rg,u}(t) = \tilde{s}_{1,\infty}^{rg,u} = -\varepsilon(1+m)/(1+M)^2$ and $\lim_{t\to\infty} \tilde{c}_1^{rg,u}(t) = \tilde{c}_{1,\infty}^{rg,u} = -\varepsilon M(1+2m-M)/(1+M)^4$. The corresponding asymptotic constants, for the two considered sets of ICs, are plotted in the figures, too. Both this unphysical outcome and the lack the 1st order contribution to the outer solution, are the most evident in the result for the substrate, in the case of the b set of ICs, corresponding to the quite large $\varepsilon = \varepsilon^b = 0.5$ ([Fig. 3.5C]).

On the other hand, the figure shows that the 1st order SPDERG approach, differently from the 0th order perturbation expansion method, is able to capture important features of the correct solutions. In detail, one can observe the presence of the three inflection points in the curves of the substrate both in [Fig. 3.5A] and in [Fig. 3.5C]. From the quantitative point of view, also the correct complex maximum values turn out to be better predicted than in the corresponding [Fig. 3.3], further confirming that the approach is particularly successful in the matching region and near it.

Therefore, on the basis of these initial results, that appear as a whole to support the SPDERG usefulness, in order both to further test its correctness and to possibly obtain better approximations to the correct solutions than the ones given by the perturbation expansion 1st order uniform approximations, we consider the 2nd order in ε, too.

3.5.2 Second order contribution

With the aim of simplifying the calculations, here we limit ourselves to study the case in which the contributions to the complex IC, at the considered orders in the expansion, in $\tau = \tau_0$, are fixed at $\tilde{c}_0^* = \tilde{c}_1^* = \tilde{c}_2^* = 0$ from the beginning, for every τ_0. This could seem unrealistic, but it is not expected to influence the result, for two reasons. On the one hand, we already showed that both $\tilde{c}_0^*$ and $\tilde{c}_1^*$ were negligible at the previous order, whereas $\tilde{c}_2^*$ should be negligible here at least in the same way as $\tilde{s}_1^*$ at the previous order. On the other hand, since one is interested in taking as initial time $\tau_0 = 0$, we expect that, as at the 1st order, the renormalization of $\tilde{s}$ implies the correct renormalization of $\tilde{c}$. Thus, it appears reasonable to set $\tilde{c}_0^* = \tilde{c}_1^* = \tilde{c}_2^* = 0$.

Once again, this last expectation can be intuitively understood, first of all, within the analogy with the matching in the perturbation expansion, where there is freedom for fixing only one of the two conditions, whereas the other turns out to be consistently satisfied, too. Moreover, also the observation that, within the SPDERG approach scheme as applied to different boundary layer problems in [27], one should be able to limit the study to one of the 2nd order equations for the inner components (here, Eq. (3.40), or Eq. (3.41)), points towards the same direction.

We anyway verify in detail that indeed one gets the same ODE to be obeyed by $\tilde{s}_1^{rg*}$ both from the study of the substrate and from the study of the complex.

Moreover, we take $\tilde{s}_2^* = 0$ from the beginning, too, since we expect that it can be anyway neglected within this approach at the 2nd order, in the same way as $\tilde{s}_1^*$ turned out to be negligible at the 1st order. Notice that we should instead allow for $\tilde{s}_2^* \neq 0$, if we were to use these solutions to calculate the 3rd order contribution, too.

Clearly, one looks for 2nd order solutions in the form:

$$\begin{cases} \tilde{s}^{rg}(\tau) = \tilde{s}_0^{rg}(\tau) + \varepsilon\tilde{s}_1^{rg}(\tau) + \varepsilon^2\tilde{s}_2^{rg}(\tau) + O(\varepsilon^3) \\ \tilde{c}^{rg}(\tau) = \tilde{c}_0^{rg}(\tau) + \varepsilon\tilde{c}_1^{rg}(\tau) + \varepsilon^2\tilde{c}_2^{rg}(\tau) + O(\varepsilon^3). \end{cases} \tag{3.58}$$

Here, $\tilde{s}_0^{rg}(\tau)$ and $\tilde{c}_0^{rg}(\tau)$ are still given by (3.43), with $\tilde{c}_0^* = 0$ in the case of the complex. Moreover, $\tilde{s}_1^{rg}(\tau)$ and $\tilde{c}_1^{rg}(\tau)$ are still given by (3.45) and (B.1), respectively, with $\tilde{c}_1^* = 0$ in the case of the complex, whereas the condition $\tilde{c}_0^* = 0$ does apply to the solution for the substrate, too.

Correspondingly, we end up to study the system:

$$\left\{ \begin{array}{rcl} \dot{\tilde{s}}_2^{rg}(\tau) & = & (\tilde{s}_0^* + m)\tilde{c}_1^{rg}(\tau) + [\tilde{c}_0^{rg}(\tau) - 1]\,\tilde{s}_1^{rg}(\tau) \\ \dot{\tilde{c}}_2^{rg}(\tau) & = & -(\tilde{s}_0^* + M)\tilde{c}_2^{rg}(\tau) - [\tilde{c}_0^{rg}(\tau) - 1]\,\tilde{s}_2^{rg}(\tau) - \tilde{s}_1^{rg}(\tau)\tilde{c}_1^{rg}(\tau), \end{array} \right. \tag{3.59}$$

whose solutions, $\tilde{s}_2^{rg}(\tau)$ and $\tilde{c}_2^{rg}(\tau)$, with ICs, at $\tau = \tau_0$, $\tilde{s}_2^{rg}(\tau_0) = 0$ and $\tilde{c}_2^{rg}(\tau_0) = 0$, are given in Appendix B.2.

The obtained formulas are anyway cumbersome, as one could expect: in fact, they contain both (implicitly) the 1st order contributions to the outer solutions and (explicitly) the 2nd order contributions to the inner ones. As anticipated, these last 2nd order contributions are calculated, in the present work, for the first time to our knowledge.

By carrying on the study as for the first order contribution, we start from the part to be renormalized of the substrate solution. The first three terms in the solution $\tilde{s}_2^{rg}(\tau)$, given by (B.3), are the secular ones, whose coefficients are reported in (B.4). The correct function, $\tilde{s}_{div}^{rg}(\tau)$, needs also to contain the constant term that appears in $s_1^{rg}(\tau)$ (given by (3.45)), since this is a leading order term at the 2nd order. In fact, it depends on the 0th order contribution to the bare substrate IC (*i.e.*, $\tilde{s}_0^*$) and it does not tend to zero for $\tau \to \infty$.

Hence, one finds:

$$\begin{array}{rcl} \tilde{s}_{div}^{rg}(\tau) & = & \tilde{s}_0^* + \varepsilon\tilde{s}_1^* - \varepsilon\dfrac{\tilde{s}_0^* + m}{(\tilde{s}_0^* + M)^2}\tilde{s}_0^* - \varepsilon\dfrac{(M-m)\tilde{s}_0^*}{\tilde{s}_0^* + M}(\tau - \tau_0) + \\ & + & \varepsilon^2\dfrac{2M(M-m)\tilde{s}_0^*}{(\tilde{s}_0^* + M)^4}\,(\tilde{s}_0^* + m)\,(\tau - \tau_0) - \varepsilon^2\dfrac{M(M-m)\tilde{s}_1^*}{(\tilde{s}_0^* + M)^2}(\tau - \tau_0) + \\ & + & \varepsilon^2\dfrac{M(M-m)^2\tilde{s}_0^*}{2\,(\tilde{s}_0^* + M)^3}(\tau - \tau_0)^2. \end{array} \tag{3.60}$$

Then, we replace the bare constants with the renormalized ones, *i.e.*:

$$\left\{ \begin{array}{rcl} \tilde{s}_0^* & = & Z_{s_0}(\tau_0, \lambda)\tilde{s}_0^{rg*}(\lambda) \\ \tilde{s}_1^* & = & Z_{s_1}(\tau_0, \lambda)\tilde{s}_1^{rg*}(\lambda), \end{array} \right. \tag{3.61}$$

with, at the 2nd order in ε:

$$\left\{ \begin{array}{rcl} Z_{s_0}(\tau_0, \lambda) & = & 1 + \varepsilon z_{s_0,1}(\tau_0, \lambda) + \varepsilon^2 z_{s_0,2}(\tau_0, \lambda), \\ Z_{s_1}(\tau_0, \lambda) & = & 1 + \varepsilon z_{s_1,1}(\tau_0, \lambda). \end{array} \right. \tag{3.62}$$

Correspondingly, we write $(\tau - \tau_0) = (\tau - \lambda) + (\lambda - \tau_0)$ in (3.60) and we absorb the secular terms in $(\lambda - \tau_0)$ by appropriately choosing the renormalization constant coefficients. In detail, $z_{s_0,1}(\tau_0, \lambda)$ is already fixed by (3.48). Nevertheless, the contributions proportional to this coefficient to the third term and to the fourth one, in (3.60) are to be taken carefully into account, since they turn out to be $O(\varepsilon^2)$.

In particular, in the case of the third term, when renormalizing $\tilde{s}_0^*$, one has:

$$-\varepsilon\frac{(\tilde{s}_0^*+m)\tilde{s}_0^*}{(\tilde{s}_0^*+M)^2} = -\varepsilon\frac{(\tilde{s}_0^{rg*}+m)\tilde{s}_0^{rg*}}{(\tilde{s}_0^{rg*}+M)^2} - \varepsilon^2\left\{ \frac{d}{d\tilde{s}_0^*}\left[\frac{(\tilde{s}_0^*+m)\tilde{s}_0^*}{(\tilde{s}_0^*+M)^2}\right]\Big|_{\tilde{s}_0^*=\tilde{s}_0^{rg*}} \right\} z_{s_0,1}(\tau_0, \lambda)s_0^{rg*}, \tag{3.63}$$

with:

$$-\varepsilon^2\left\{ \frac{d}{d\tilde{s}_0^*}\left[\frac{(\tilde{s}_0^*+m)\tilde{s}_0^*}{(\tilde{s}_0^*+M)^2}\right]\Big|_{\tilde{s}_0^*=\tilde{s}_0^{rg*}} \right\} z_{s_0,1}(\tau_0, \lambda)s_0^{rg*} =$$

$$
\begin{aligned}
&= -\varepsilon^2 \left[\frac{(M-m)\tilde{s}_0^{rg*}}{(\tilde{s}_0^{rg*}+M)^3} + \frac{M(\tilde{s}_0^{rg*}+m)}{(\tilde{s}_0^{rg*}+M)^3} \right] z_{s_0,1}(\tau_0,\lambda)\tilde{s}_0^{rg*} = \\
&= -\varepsilon^2 \left[\frac{(M-m)^2(\tilde{s}_0^{rg*})^2}{(\tilde{s}_0^{rg*}+M)^4} + \frac{M(M-m)\tilde{s}_0^{rg*}}{(\tilde{s}_0^{rg*}+M)^4}(\tilde{s}_0^{rg*}+m) \right] (\lambda-\tau_0). \qquad (3.64)
\end{aligned}
$$

In fact, both of these two terms contribute and they have to be taken into account in the definition of $z_{s_0,2}(\tau_0,\lambda)$. Let us note, moreover, that $d[\tilde{s}_0^*/(\tilde{s}_0^*+M)]/d\tilde{s}_0^* = M/(\tilde{s}_0^*+M)^2$. Thus, one can check that the 2nd order contribution proportional to the same coefficient, $z_{s_0,1}(\tau_0,\lambda)$ (that originates from the fourth term in $\tilde{s}_{div}^{rg}$, in the same way as we just explained in detail with the formulas (3.63) and (3.64)), partially absorbs the term proportional to $(\tau-\lambda)(\lambda-\tau_0)$ (that originates from the last term) and partially contributes to $z_{s_0,2}(\tau_0,\lambda)$.

Thus, by taking:

$$
\left\{
\begin{aligned}
z_{s_0,2}(\tau_0,\lambda) &= \left[\frac{(M-m)^2\tilde{s}_0^{rg*}}{(\tilde{s}_0^{rg*}+M)^4} - \frac{M(M-m)}{(\tilde{s}_0^{rg*}+M)^4}(\tilde{s}_0^{rg*}+m) \right] (\lambda-\tau_0) + \\
&+ \frac{M(M-m)^2}{2(\tilde{s}_0^{rg*}+M)^3}(\lambda-\tau_0)^2 \qquad (3.65) \\
z_{s_1,1}(\tau_0,\lambda) &= \frac{M(M-m)}{(\tilde{s}_0^{rg*}+M)^2}(\lambda-\tau_0),
\end{aligned}
\right.
$$

we end up with an expression for $\tilde{s}_{div}^{rg}(\tau,\lambda)$ that is exactly the same as the one given by (3.60), with $\tilde{s}_0^* \to \tilde{s}_0^{rg*}(\lambda)$, $\tilde{s}_1^* \to \tilde{s}_1^{rg*}(\lambda)$ and $\tau_0 \to \lambda$. Moreover, though we wrote explicitly the first two terms in $z_{s_0,2}$ to make evident the different origins of the contributions, their sum can be simplified in such a way that the coefficient of the term proportional to $(\lambda-\tau_0)$ is simply equal to $-m(M-m)/(\tilde{s}_0^{rg*}+M)^3$.

Therefore, we avoid to write explicitly $\tilde{s}_{div}^{rg}(\tau,\lambda)$ and we study directly its derivative with respect to λ. We obtain, at the 2nd order:

$$
\begin{aligned}
\frac{d\tilde{s}_{div}^{rg}(\tau,\lambda)}{d\lambda} &= \frac{d\tilde{s}_0^{rg*}(\lambda)}{d\lambda} + \varepsilon\frac{d\tilde{s}_1^{rg*}(\lambda)}{d\lambda} + \varepsilon\frac{(M-m)\tilde{s}_0^{rg*}(\lambda)}{\tilde{s}_0^{rg*}(\lambda)+M} + \\
&+ \varepsilon^2\frac{(M-m)^2\left[\tilde{s}_0^{rg*}(\lambda)\right]^2}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^4} + \varepsilon^2\frac{M(M-m)\tilde{s}_0^{rg*}(\lambda)}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^4}[\tilde{s}_0^{rg*}(\lambda)+m] + \\
&+ \varepsilon^2\frac{M(M-m)^2\tilde{s}_0^{rg*}(\lambda)}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^3}(\tau-\lambda) - \varepsilon^2\frac{2M(M-m)\tilde{s}_0^{rg*}(\lambda)}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^4}\left[\tilde{s}_0^{rg*}(\lambda)+m\right] + \\
&+ \varepsilon^2\frac{M(M-m)\tilde{s}_1^{rg*}(\lambda)}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^2} - \varepsilon^2\frac{M(M-m)^2\tilde{s}_0^{rg*}(\lambda)}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^3}(\tau-\lambda). \qquad (3.66)
\end{aligned}
$$

In fact, one already makes use of the known 1st order result on $d\tilde{s}_0^{rg*}/d\lambda$, given by Eqs. (3.51), in the derivation of this equation.

Here, for the sake of clarity, we also wrote explicitly both of the 2nd order terms proportional to $(\tau-\lambda)$, that obviously cancel each other. In fact, this appears a quite consistent result of the present approach, since they have completely different origins. Indeed, the first of these terms originates from part of the derivative with respect to λ of the 1st order term, proportional to $(\tau-\lambda)$, in $\tilde{s}_{div}^{rg}(\tau,\lambda)$ (*i.e.*, from the term that is equal to $-\varepsilon(M-m)\{d[\tilde{s}_0^{rg*}/(\tilde{s}_0^{rg*}+M)]/d\lambda\}(\tau-\lambda)$). The second of these terms originates, instead, from the derivative with respect to λ of the last term in $\tilde{s}_{div}^{rg}(\tau,\lambda)$, that is proportional to $(\tau-\lambda)^2$.

Interestingly, the second term proportional to ε^2 (part of the contribution that originates from the constant term in $\tilde{s}_1^{rg}(\tau)$), letting aside a factor 2, is the same as the one that is found when deriving with respect to λ the first of the terms proportional to $\varepsilon^2(\tau-\lambda)$ in (3.60), *i.e.*, the term in $\tilde{s}_{div}^{rg}(\tau,\lambda)$ that corresponds to the fourth term proportional to ε^2 here, but it has opposite sign.

In fact, both of these results are obtained in a similar way to the one that we previously described in detail, in the context of the derivation of the appropriate renormalization constants. Indeed, from Eqs. (3.51), one has $(\lambda-\tau_0)d\tilde{s}_0^{rg*}/d\lambda = -\varepsilon z_{s_0,1}(\tau_0,\lambda)\tilde{s}_0^{rg*}$.

Finally, by imposing the scaling condition $d\tilde{s}_{div}^{rg}(\tau,\lambda)/d\lambda = 0$ and by correctly grouping the terms in ε and in ε^2, we derive the two ODEs to be obeyed by $\tilde{s}_0^{rg*}(\lambda)$ and $\tilde{s}_1^{rg*}(\lambda)$, respectively:

$$\left\{\begin{array}{rcl} \dfrac{d\tilde{s}_0^{rg*}(\lambda)}{d\lambda} & = & -\varepsilon\dfrac{(M-m)\tilde{s}_0^{rg*}(\lambda)}{\tilde{s}_0^{rg*}(\lambda)+M} \\ \dfrac{d\tilde{s}_1^{rg*}(\lambda)}{d\lambda} & = & \varepsilon\left\{-\dfrac{(M-m)^2\left[\tilde{s}_0^{rg*}(\lambda)\right]^2}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^4}+\dfrac{M(M-m)\tilde{s}_0^{rg*}(\lambda)}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^4}\left[\tilde{s}_0^{rg*}(\lambda)+m\right]+\right. \\ & - & \left.\dfrac{M(M-m)\tilde{s}_1^{rg*}(\lambda)}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^2}\right\}= \\ & = & \varepsilon\left\{\dfrac{m(M-m)\tilde{s}_0^{rg*}(\lambda)}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^3}-\dfrac{M(M-m)\tilde{s}_1^{rg*}(\lambda)}{\left[\tilde{s}_0^{rg*}(\lambda)+M\right]^2}\right\}. \end{array}\right. \tag{3.67}$$

The first of these ODEs is just the already known 1st order result for $d\tilde{s}_0^{rg*}(\lambda)/d\lambda$ reported in Eqs. (3.51). Actually, at the 2nd order, the interesting ODE is the one to be obeyed by $\tilde{s}_1^{rg*}(\lambda)$.

Let us remark that, when making the final transformation $\lambda\to\tau=t/\varepsilon$ and when recalling that $\tilde{s}_0^{rg*}(t)=\tilde{s}_0^{out}(t)$ (given by (3.12)), the ODE to be obeyed by $\tilde{s}_1^{rg*}(t)$ turns out to be different from the one for the 1st order outer substrate within the perturbation expansion method, reported in Eqs. (3.22). Here, we wrote the equation both with and without the simplification, that is due to the sum of the first two terms, just in order to make evident the difference, since it is the first term in the not simplified expression that was absent there. Indeed, to get in particular the same coefficient as in the ODE in Eqs. (3.22), for the second term in the present not simplified expression, it is essential to correctly take into account the constant in $\tilde{s}_1^{rg}(\tau)$ in the part to be renormalized of the function (at the 2nd order). In fact, as outlined, the contribution of the only term proportional to ε^2 would give an incorrect (twice larger) coefficient.

Most interestingly, the present ODE turns out to be simpler than the one encountered in the perturbation expansion method. One can check that the solution is given by:

$$\tilde{s}_1^{rg*}(t)=\frac{m\tilde{s}_0^{out}(t)}{M\left[\tilde{s}_0^{out}(t)+M\right]}\log\left[\frac{\tilde{s}_0^{out}(t)+M}{(1+M)\tilde{s}_0^{out}(t)}\right], \tag{3.68}$$

with the choice $\tilde{s}_1^{rg*}(0)=0$, that is reasonable in this context, since it allows to get, correctly, $\tilde{s}_2^{rg,u}(0)=1$. It is also important to stress that the solution also satisfies the asymptotic condition $\lim_{t\to\infty}\tilde{s}_1^{rg*}(t)=0$, though this in fact implies $\lim_{t\to\infty}\tilde{s}_2^{rg,u}(t)=O(\varepsilon^2)$.

On the other hand, when adding the 1st order term that originates from the replacement, $\tilde{s}_0^*\to\tilde{s}_0^{rg,*}(\lambda)$, in the 1st order constant term, in the part to be renormalized of the function, the total 1st order *outer* contribution to the SPDERG 2nd order uniform approximation, in the case of the substrate, is just:

$$\tilde{s}_1^{rg,out}(t)=-\frac{[\tilde{s}_0^{out}(t)+m]\tilde{s}_0^{out}(t)}{[\tilde{s}_0^{out}(t)+M]^2}+\tilde{s}_1^{rg*}(t)=\tilde{s}_1^{out}(t). \tag{3.69}$$

Thus, one recovers the 1st order outer contribution to the perturbation expansion uniform approximation reported in (3.23). Indeed, this was not a result to be taken for granted [27]. In particular, the result implies that $\tilde{s}_1^{rg,out}(0)=-(1+m)/(1+M)^2$.

Hence, we recall that $\mathcal{R}_2^s(\tau)$ is to be evaluated from the remaining part of the solution in $\tilde{s}_0^*=1$, $\tilde{s}_1^*=0$ and $\tau_0=0$. It is given by (see (B.3) and (B.5), in Appendix B.2):

$$\mathcal{R}_2^s(\tau)=D_{s_2^{rg}}(1)+\left[F_{s_2^{rg}}(1)+H_{s_2^{rg}}(1)\tau+J_{s_2^{rg}}(1)\tau^2\right]e^{-(1+M)\tau}+K_{s_2^{rg}}(1)e^{-2(1+M)\tau}=$$

$$
\begin{aligned}
=&-\frac{1}{2(1+M)^5}[2M^2(2m+1)-M(6m^2+5m+3)-m^2+m]+\\
&+\frac{1}{(1+M)^5}[M^2(2m+1)-M(3m^2+3m+2)-m^2-m-1]e^{-(1+M)\tau}+\\
&-\left\{\frac{1}{(1+M)^4}[M^2+M(m^2+1)-2m-1]\tau-\frac{(M-m)}{2(1+M)^3}(m+1)\tau^2\right\}e^{-(1+M)\tau}+\\
&+\frac{(m+1)}{2(1+M)^5}(M+m+2)e^{-2(1+M)\tau}. \qquad (3.70)
\end{aligned}
$$

Here we used, in particular, the coefficient values reported in (B.5).

Correspondingly, we get the complete result for the SPDERG 2nd order uniform approximation to the dimensionless substrate solution:

$$
\tilde{s}_2^{rg,u}(t)=\tilde{s}_1^{rg,u}(t)-\varepsilon\left\{\frac{\tilde{s}_0^{out}(t)+m}{[\tilde{s}_0^{out}(t)+M]^2}\tilde{s}_0^{out}(t)-\frac{1+m}{(1+M)^2}-\tilde{s}_1^{rg*}(t)\right\}+\varepsilon^2\mathcal{R}_2^s(t/\varepsilon)+O(\varepsilon^2). \qquad (3.71)
$$

Here: $\tilde{s}_1^{rg,u}(t)$ is given by (3.57); $\tilde{s}_0^{out}(t)=\tilde{s}_0^{rg*}(t)$ is given by (3.12); $\tilde{s}_1^{rg*}(t)$ and $\mathcal{R}_2^s(t/\varepsilon)$ are given by (3.68) and (3.70), respectively. We wrote the solution in the present form with the aim of making as evident as possible analogies and differences with the 1st order perturbation expansion result. In particular, the constant term that is equal to $-(1+m)/(1+M)^2$ originates, here, from the constant term depending on $\tilde{s}_0^*$ that appeared (calculated in $\tilde{s}_0^*=1$) in $\tilde{s}_1^{rg,u}(t)$, the one that has been now included in $\tilde{s}_{div}^{rg}$, at the 2nd order. Moreover, one can notice that the result implies, as expected, $\lim_{t\to\infty}\tilde{s}_2^{rg,u}(t)=\tilde{s}_{2,\infty}^{rg,u}=\varepsilon^2 D_{s_2^{rg}}(1)=O(\varepsilon^2)$, where $D_{s_2^{rg}}(1)$ is the first of the terms in (3.70), reported in Appendix B.2.

For the complex, the 2nd order terms to be added to the part of the solution to be renormalized are the three ones that appear in $\tilde{c}_{2,div}^{rg}(\tau)$, that is reported in (B.7), in Appendix B.2. Moreover, we have again to take into account also the constant term in $\tilde{c}_1^{rg}(\tau)$, *i.e.*, the one that corresponds to the coefficient $B_{c_1^{rg}}(\tilde{s}_0^*,\tilde{s}_1^*,\tilde{c}_0^*)$, that is reported in (B.2), in Appendix B.2, and that needs to be calculated in $\tilde{c}_0^*=0$. We obtain:

$$
\begin{aligned}
\tilde{c}_{div}^{rg}(\tau) &= \frac{\tilde{s}_0^*}{\tilde{s}_0^*+M}-\varepsilon\frac{M(\tilde{s}_0^*-M+2m)}{(\tilde{s}_0^*+M)^4}\tilde{s}_0^*+\varepsilon\frac{M\tilde{s}_1^*}{(\tilde{s}_0^*+M)^2}-\varepsilon\frac{M(M-m)\tilde{s}_0^*}{(\tilde{s}_0^*+M)^3}(\tau-\tau_0)+\\
&+ \varepsilon^2 A_{c_{2,div}^{rg}}(\tilde{s}_0^*)(\tau-\tau_0)+\varepsilon^2 B_{c_{2,div}^{rg}}(\tilde{s}_0^*,\tilde{s}_1^*)(\tau-\tau_0)+\varepsilon^2 C_{c_{2,div}^{rg}}(\tilde{s}_0^*)(\tau-\tau_0)^2, \qquad (3.72)
\end{aligned}
$$

Here, the coefficients $A_{c_{2,div}^{rg}}$, $B_{c_{2,div}^{rg}}$ and $C_{c_{2,div}^{rg}}$, of the three 2nd order secular terms, are given by (B.8), once again in Appendix B.2.

The proof, that is reported in Appendix B.3 (see Eqs. (B.14)-(B.17) and the following discussion), involves quite complicated calculations. Nevertheless, one can check that here one obtains a function $\tilde{c}_{div}^{rg}(\tau,\lambda)$ exactly corresponding to (3.72), with $\tilde{s}_0^*\to\tilde{s}_0^{rg*}(\lambda)$, $\tilde{s}_1^*\to\tilde{s}_1^{rg*}(\lambda)$ and $\tau_0\to\lambda$, in the same way as we showed in detail in the study of the substrate, with the same renormalization coefficients of the bare substrate IC ($z_{s_0,1}$ already fixed by (3.48); $z_{s_0,2}$ and $z_{s_1,1}$ already fixed by (3.65), respectively). This was in fact the expected result, both from the point of view of the analogy with the matching in the perturbation expansion and from the observation that it should have been alternatively possible to limit the study to Eq. (3.40), or Eq. (3.41).

Moreover, as expected within the same context, we verify that, when imposing the scaling condition, $d\tilde{c}_{div}^{rg}(\tau,\lambda)/d\lambda=0$, one recovers, once again, the two ODEs given by (3.67), to be obeyed by $\tilde{s}_0^{rg*}(\lambda)$ and $\tilde{s}_1^{rg*}(\lambda)$. Thus, one obtains, in particular, the same result on $\tilde{s}_1^{rg*}(\lambda)$ (as reported in Appendix B.3, too, see Eq. (B.20)).

In detail, also in the case of the complex, one needs to use the known 1st order result on $d\tilde{s}_0^{rg*}(\lambda)/d\lambda$, in the derivation of the 2nd order one on $d\tilde{s}_1^{rg*}(\lambda)/d\lambda$. Also here, the result on the

derivative with respect to λ is obtained in a very similar way to the proof of the correspondence between $\tilde{c}^{rg}_{div}(\tau,\lambda)$ and the original $\tilde{c}^{rg}_{div}(\tau)$, given by (3.72).

Clearly, in the present case, there is a definitely larger number of relevant terms that anyway either cancel each other or contribute in the correct way to the final result. Therefore, the verification, reported in Appendix B.3, of the correctness of the expectations we made, appears to give further consistency to the whole approach.

Correspondingly, we obtain a 1st order *outer* contribution to the SPDERG 2nd order uniform approximation, for the dimensionless complex concentration, that turns out to be exactly equal to the 1st order perturbation expansion outer contribution, given by (3.24). It is obtained, here, from an algebraic relation, that could appear different from the one reported in Eqs. (3.22), but that is, in fact, equivalent. This becomes evident when writing $\tilde{s}^{rg*}_1(t)$ in terms of $\tilde{s}^{out}_0(t)$ and $\tilde{s}^{out}_1(t)$, by means of (3.69):

$$\tilde{c}^{rg,out}_1(t) = -\frac{M(\tilde{s}^{rg*}_0(t) - M + 2m)}{\left[\tilde{s}^{rg*}_0(t) + M\right]^4}\tilde{s}^{rg*}_0(t) + \frac{M\tilde{s}^{rg*}_1(t)}{\left[\tilde{s}^{rg*}_0(t) + M\right]^2} = \tilde{c}^{out}_1(t). \tag{3.73}$$

The equality can also be checked by reminding that $\tilde{s}^{rg*}_0(t) = \tilde{s}^{out}_0(t)$ and by using the known result for $\tilde{s}^{rg*}_1(t)$, reported in (3.68).

The SPDERG 2nd order uniform approximation for the complex is then obtainable by taking into account also the remaining part of the 2nd order inner solution. It is given by:

$$\begin{aligned}\tilde{c}^{rg,u}_2(t) = \tilde{c}^{rg,u}_1(t) - \varepsilon\left\{\frac{M(\tilde{s}^{rg*}_0(t) - M + 2m)}{\left[\tilde{s}^{rg*}_0(t) - M\right]^4}\tilde{s}^{rg*}_0(t) - \frac{M\tilde{s}^{rg*}_1(t)}{\left[\tilde{s}^{rg*}_0(t) + M\right]^2} - \frac{M(1 - M + 2m)}{(1+M)^4}\right\} + \\ + \varepsilon^2\mathcal{R}^c_2(t/\varepsilon) + O(\varepsilon^2).\end{aligned} \tag{3.74}$$

Here, we attempt to outline once more similarities and differences with the perturbation expansion result. In fact, because of (3.73), one can equivalently write the terms in curly brackets as $\tilde{c}^{out}_1(t)$ minus the constant terms that in the standard perturbation expansion method appear twice.

In detail, the various terms that appear in (3.74) are reported: $\tilde{c}^{rg,u}_1(t)$ in (3.57); $\tilde{s}^{rg*}_0(t) = \tilde{s}^{out}_0(t)$ in (3.12); $\tilde{s}^{rg*}_1(t)$ in (3.68); and $\mathcal{R}^c_2(\tau)$, correctly evaluated in $\tau_0 = 0$, $\tilde{s}^*_0 = 1$ and $\tilde{s}^*_1 = 0$ in (B.9), with the coefficients given by (B.11), in Appendix B.3.

This uniform approximation verifies the IC $\tilde{c}^{rg,u}_2(0) = 0$, whereas one finds:

$$\lim_{t\to\infty}\tilde{c}^{rg,u}_2(t) = \tilde{c}^{rg,u}_{2,\infty} = \varepsilon^2 A_{\mathcal{R}^c_2}(1,0) = O(\varepsilon^2)\ . \tag{3.75}$$

Here, $A_{\mathcal{R}^c_2}(1,0)$ is the constant term in $\mathcal{R}^c_2(\tau)$, whose detailed dependence on $\tilde{s}^*_0$ and $\tilde{s}^*_1$ is reported in (B.10), and that is calculated in $\tilde{s}^*_0 = 1$, $\tilde{s}^*_1 = 0$ in (B.11), in Appendix B.3.

We plot in [Fig. 3.6] our numerical results on the SPDERG 2nd order uniform approximations for the substrate and the complex, respectively, for the two considered sets of ICs. The plots are, as usual, in comparison with the numerical solutions of the original problem, given by Eqs. (3.6) (the same curves as in the previous figures).

Since they contain the 2nd order terms of the inner solutions (this is indeed the only difference), these are definitely better approximations than the 1st order perturbation expansion uniform approximations, in a region that encompasses the matching one. Indeed, the results here are nearly indistinguishable, within our numerical precision, from the correct ones on a definitely larger time window. This is true also in the particularly unfavourable case of the substrate in [Fig. 3.6C] and the outcome is clearly different from the one observed in the same case at the 1st order, when applying the standard perturbation expansion method, that is reported in [Fig. 3.4C].

Nevertheless, one can still note a minor discrepancy at large times, that is at least partially to be related to the failure of the approximations in reproducing the asymptotically vanishing solutions. Actually, on the basis of the results that we already obtained in the present study, this failure seems correctable in a reasonable way. In the following Subsection 3.5.3, we just introduce refined

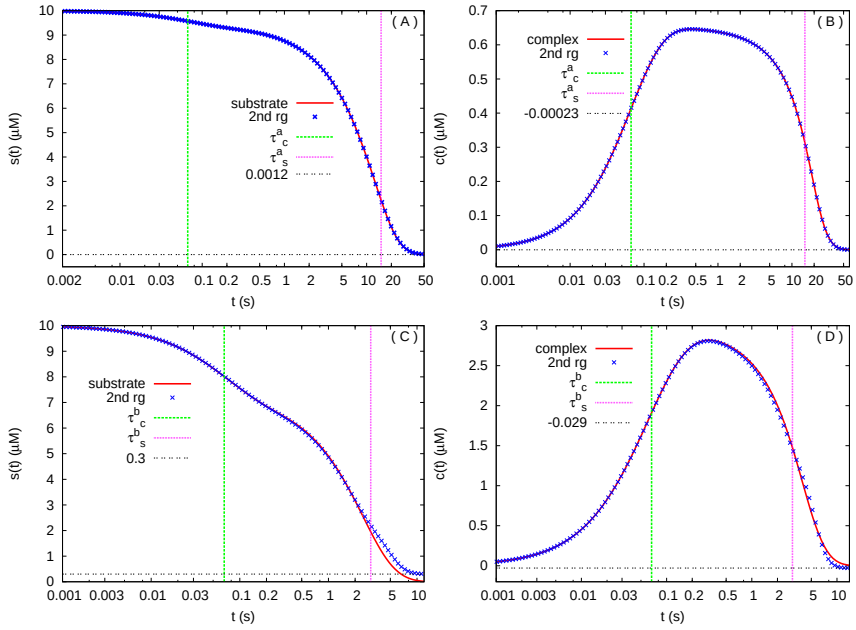

Figure 3.6: **Comparison between the numerical solutions and the 2nd order SPDERG uniform approximations, for the two considered sets of ICs.** In A) and C) we present the behaviour of the concentrations of the substrate $s(t)$, whereas in B) and D) we present the one of the concentrations of the complex $c(t)$ for the a and b sets of ICs given by (3.8), respectively. Hence, in A) and B) we are in the case with $\varepsilon = \varepsilon^a = 0.1$, whereas in C) and D) we are in the one with $\varepsilon = \varepsilon^b = 0.5$. We plot both the numerical solutions of Eqs. (3.6), already shown in the previous figures, and the analytical solutions computed from the SPDERG 2nd order uniform approximations (with a standard numerical approximation to the Lambert function), as given by (3.71) and (3.74), respectively. We plot moreover the (physically meaningless) asymptotic limits of these analytical solutions: $s_0^a \tilde{s}^{rg,u}_{2,\infty} \simeq 0.0012$ (μM), $e_0^a \tilde{c}^{rg,u}_{2,\infty} \simeq -0.00023$ (μM) and $s_0^b \tilde{s}^{rg,u}_{2,\infty} \simeq 0.3$ (μM), $e_0^b \tilde{c}^{rg,u}_{2,\infty} \simeq -0.029$ (μM), for the a and b sets of ICs, respectively. We finally plot our corresponding rough evaluations of the two different time scales involved, too, with τ_s describing the substrate decay time and τ_c the complex saturation time. Notice that the time is in logarithmic scale.

SPDERG 2nd order uniform approximations that satisfy the asymptotic conditions $\lim_{t\to\infty} s(t) = \lim_{t\to\infty} c(t) = 0$, too.

3.5.3 Refined SPDERG second order uniform approximations

Within the framework of the already obtained results, by assuming that the found solution behaviours could be iterated, one can hypothesize that, in the actual SPDERG approach, the constant term at a given order will contribute to the outer component of the solution at the following one. Actually, this appears to us the SPDERG approach ingredient that is equivalent to take into account both a part of the matching conditions and the removing of one of the constants that otherwise would appear twice in the uniform approximations of the standard perturbation expansion method, though the observation needs to be better formalized for investigating its possible generalizations.

In fact, in the present case, when passing from the 2nd to the 3rd order, these iteratively expected solution behaviours should be obtainable by means of the substitutions $\tilde{s}_0^* \to \tilde{s}_0^{rg*}(t)$ and $\tilde{s}_1^* \to \tilde{s}_1^{rg*}(t)$ in the constant terms $D_{s_2^{rg}}(\tilde{s}_0^*) + E_{s_2^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*)$ and $A_{\mathcal{R}_2^c}(\tilde{s}_0^*, \tilde{s}_1^*)$. These terms appear in the 2nd order inner solution for the substrate and the complex, respectively, that are given by (B.5) and (B.10), in Appendix B.2. We underline that, despite the substitutions, they remain terms of 2nd order, in agreement with the general consideration that, at the n-th order, within the SPDERG approach, one in fact obtains the $(n-1)$-th order outer components of the corresponding perturbation expansion uniform approximations, *i.e.*, their leading order terms [26, 27].

Therefore, we can finally consider as refined SPDERG 2nd order uniform approximations the functions $\tilde{s}_{2,r}^{rg,u}(t)$ and $\tilde{c}_{2,r}^{rg,u}(t)$, that satisfy by construction the physically correct asymptotic conditions $\lim_{t\to\infty} \tilde{s}_{2,r}^{rg,u}(t) = \lim_{t\to\infty} \tilde{c}_{2,r}^{rg,u}(t) = 0$. These are given by:

$$\begin{cases} \tilde{s}_{2,r}^{rg,u}(t) = \tilde{s}_2^{rg,u}(t) + \varepsilon^2 \left\{ D_{s_2^{rg}}[\tilde{s}_0^{rg*}(t)] + E_{s_2^{rg}}[\tilde{s}_0^{rg*}(t), \tilde{s}_1^{rg*}(t)] - D_{s_2^{rg}}(1) \right\} + O(\varepsilon^2) \\ \tilde{c}_{2,r}^{rg,u}(t) = \tilde{c}_2^{rg,u}(t) + \varepsilon^2 \left\{ A_{\mathcal{R}_2^c}[\tilde{s}_0^{rg*}(t), \tilde{s}_1^{rg*}(t)] - A_{\mathcal{R}_2^c}(1,0) \right\} + O(\varepsilon^2), \end{cases} \tag{3.76}$$

We remind that $E_{s_2^{rg}}(1,0) = 0$, whereas: $\tilde{s}_2^{rg,u}(t)$ and $\tilde{c}_2^{rg,u}(t)$ are reported in (3.71) and (3.74), respectively; $\tilde{s}_0^{rg*}(t) = \tilde{s}_0^{out}(t)$ in (3.23); $\tilde{s}_1^{rg*}(t)$ in (3.68).

It is not to be taken for granted that these approximations could turn out to be better than the previously considered ones, since they, anyway, lack a part of the 2nd order outer contributions. On the other hand, these appear to us the most refined SPDERG 2nd order uniform approximations that one can propose, by exploiting as much as possible the obtained results.

We present the corresponding substrate and complex behaviours, as usual for the two different considered sets of ICs, in [Fig. 3.7]. The plots are, once more, in comparison with the numerical solutions of the original problem, given by Eqs. (3.6) (the same curves as in the previous figures), too.

In fact, one could already observe in the previous [Fig. 3.6A], [Fig. 3.6B] that there was no detectable difference between the SPDERG 2nd order uniform approximations and the correct solutions in the case of the a set of ICs. Indeed, this set corresponds to the relatively small $\varepsilon = \varepsilon^a = 0.1$ and the perturbation expansion 1st order uniform approximations showed no detectable difference with respect to the correct solutions for this set of ICs ([Fig. 3.4A], [Fig. 3.4B]), too. In this case, we limit ourselves to underline that the here proposed SPDERG 2nd order approximations, that are presented in [Fig. 3.7A] for the substrate concentration and in [Fig. 3.7B] for the complex one, respectively, are, moreover, also rigorously asymptotically vanishing. This implies, in particular, that they are nearly indistinguishable from the correct solutions, within our numerical precision, as much as the perturbation expansion 1st order results.

On the other hand, when looking at [Fig. 3.7C] and [Fig. 3.7D], that show the behaviour of the substrate and complex concentration, respectively, for the b case of ICs, *i.e.*, for the quite large value of the expansion parameter $\varepsilon = \varepsilon^b = 0.5$, the plots turn out to be not enough detailed for

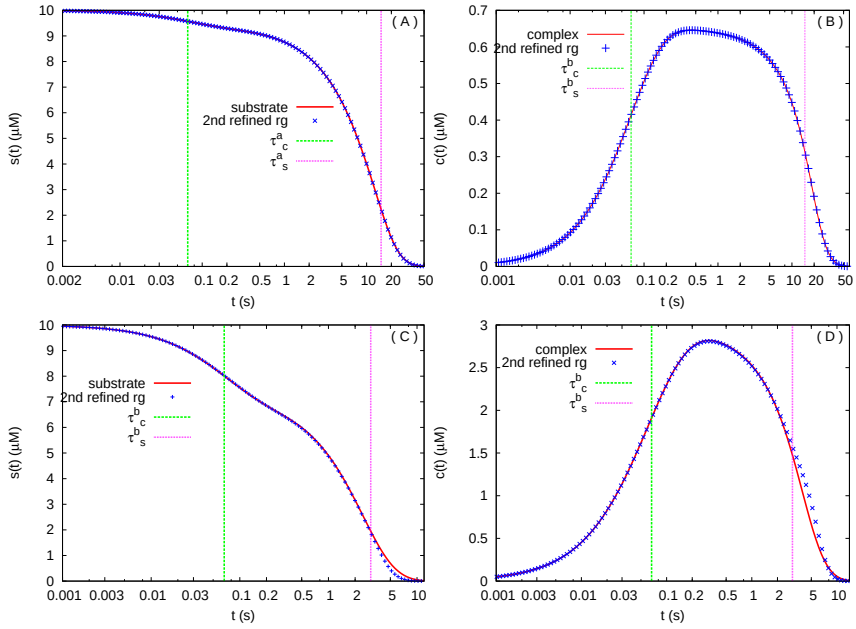

Figure 3.7: **Comparison between the numerical solutions and the refined 2nd order SPDERG uniform approximations, for the two considered sets of ICs.** In A) and C) we present the behaviour of the concentrations of the substrate $s(t)$, whereas in B) and D) we present the one of the concentrations of the complex $c(t)$ for the a and b sets of ICs given by (3.8), respectively. Hence, in A) and B) we are in the case with $\varepsilon = \varepsilon^a = 0.1$, whereas in C) and D) we are in the one with $\varepsilon = \varepsilon^b = 0.5$. We plot both the numerical solutions of Eqs. (3.6), already shown in the previous figures, and the analytical solutions computed from the refined SPDERG 2nd order uniform approximations (with a standard numerical approximation to the Lambert function), as given by (3.76). We finally plot our corresponding rough evaluations of the two different time scales involved, too, with τ_s describing the substrate decay time and τ_c the complex saturation time. Notice that the time is in logarithmic scale.

making evident the differences between the present approximated solutions and the ones without the refinement. Both for this reason and for roughly quantifying our various qualitative observations, we are led to a more careful study.

3.5.4 A conclusive comparison

We show in [Fig. 3.8] the detailed time depending behaviours of the substrate concentration, $s(t)$, and of the complex one, $c(t)$, for the case that corresponds to the quite large value of the expansion parameter $\varepsilon = \varepsilon^b = 0.5$. First of all, we plot separately and differently the two relevant parts of the time window, *i.e.*, the time is in logarithmic scale only in the first part. Here, we compare the different *best* approximations that we both recall and obtain in the present work: i) the perturbation expansion 1st order uniform approximations (as given by (3.28), already presented in [Fig. 3.4C], [Fig. 3.4D]); ii) the SPDERG 2nd order uniform approximations (as given by (3.71), (3.74), already presented in [Fig. 3.6C], [Fig. 3.6D]); iii) the refined SPDERG 2nd order uniform approximations (as given by (3.76), already presented in [Fig. 3.7C], [Fig. 3.7D]).

In detail, we neglect the initial time interval, up to $t = 0.03s$ for the substrate and to $t = 0.08s$ for the complex, respectively. Indeed, in this interval, the different results are indistinguishable, within our numerical precision, both each other and with the correct numerical solutions. Instead, we show with higher definition than previously, as usual in logarithmic time scale, the central intervals, *i.e.*, the ones that encompass the matching region. These intervals are $t \sim 0.03 - 3.5\ s$ for the substrate ([Fig. 3.8A]) and $t \sim 0.08 - 2.5\ s$ for the complex ([Fig. 3.8C]), respectively. Finally, we present

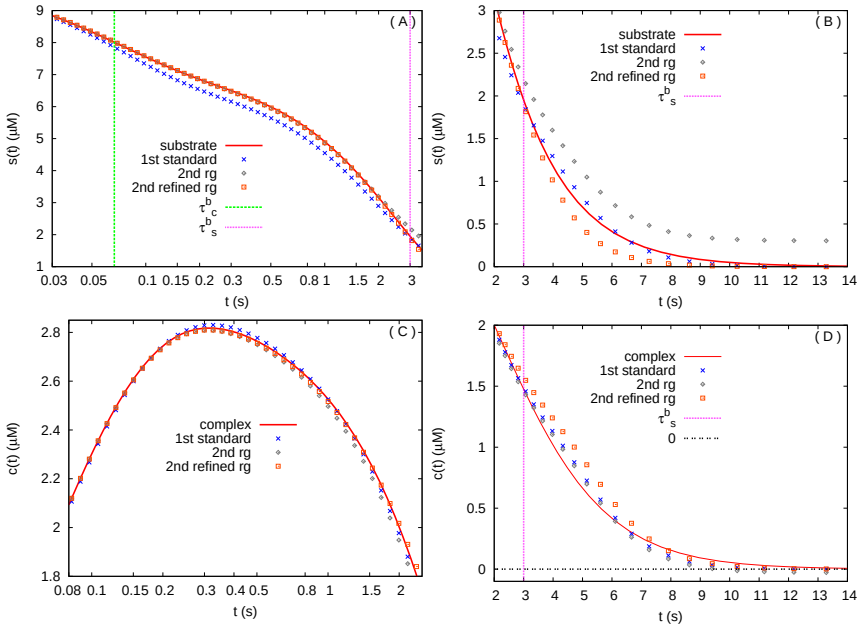

Figure 3.8: **Comparison between the numerical solutions, the 1st order perturbation expansion uniform approximations, the 2nd order SPDERG uniform approximations and the refined 2nd order SPDERG uniform approximations, for the more unfavourable of the two considered sets of ICs.** In A) (respectively C)) we present the behaviour of the concentration of the substrate $s(t)$ (respectively of the complex $c(t)$) in the central part of the time window, in logarithmic time scale, whereas in B) (respectively D)) we present the large time behaviour of these two quantities. The results are for the b set of ICs given by (3.8), *i.e.*, the one that corresponds to $\varepsilon = \varepsilon^b = 0.5$. We plot the numerical solutions of Eqs. (3.6), already shown in the previous figures, and we compare with these correct behaviours: i) the perturbation expansion 1st order uniform approximations (as given by (3.28), already presented in [Fig. 3.4C], [Fig. 3.4D]); ii) the SPDERG 2nd order uniform approximations (as given by (3.71), (3.74), already presented in [Fig. 3.6C], [Fig. 3.6D]); iii) the refined SPDERG 2nd order uniform approximations (as given by (3.76), already presented in [Fig. 3.7C], [Fig. 3.7D]). All these uniform approximations are computed by means of the same standard numerical approximation to the Lambert function. When they belong to the time window, we finally plot our corresponding rough evaluations of the two different time scales involved, too, with τ_s^b describing the substrate decay time and τ_c^b the complex saturation time.

in non logarithmic time scale the relevant large time window, [Fig. 3.8B] and [Fig. 3.8D], for the substrate and for the complex, respectively. In fact, this window ranges up to $t \sim 12s$ for the substrate and up to $t \sim 14s$ for the complex, since at longer times (as it is clear from [Fig. 3.8B], in particular) the solutions have already reached, within our numerical precision, their asymptotic values (that are different from zero in the case of the non refined SPDERG).

Once more, the curves shown in [Fig. 3.8A] make evident that, in the case of the substrate, both of the present SPDERG 2nd order uniform approximations are definitely better approximations to the correct solutions than the perturbation expansion 1st order ones, both in the matching region and near it. They, moreover, allow to naked-eye evaluate the range in which this happens. In fact, this range corresponds to $t \sim 0.05 - 2\ s$, hence it covers about the 15% of the whole relevant time window, in the case without refinement. Instead, this range corresponds to $t \sim 0.05 - 3\ s$ (the longer time is of the same order of τ_s^b), hence it covers about the 25% of the whole relevant time window, in the case with the refinement. On the other hand, for $t \gtrsim 3s$ ([Fig. 3.8B]), the refined SPDERG 2nd order uniform approximation appears to tend to zero slightly too rapidly, with respect both to the correct solution and to the perturbation expansion 1st order result, though it is correctly asymptotically vanishing and, moreover, it makes a smaller error than the same SPDERG approximation without the refinement.

In the case of the complex, as already anticipated, the curves shown in [Fig. 3.8C] make clear that the perturbation expansion 1st order uniform approximation slightly over-evaluates the value of its maximum. In fact, it appears to over-evaluate the complex dynamical behaviour in the whole range $t \sim 0.25 - 0.7\ s$, that is a time interval roughly centered around the maximum abscissa (in logarithmic time scale). On the other hand, both of the SPDERG 2nd order uniform approximations are successful in correctly capturing the maximum value of the complex numerical solution. In particular, the refined SPDERG approximation turns out to be the one more in agreement with the correct result up to the larger time $t \sim 2s$, *i.e.*, in about the 15% of the whole relevant time window. It is moreover both as much correct as the 1st order result of the standard method (though going to zero slightly more rapidly than the correct solution) and definitely better than the same approximation without refinement at large times. In particular, this is true for $t \sim 9 - 14\ s$, that covers about a remaining 35% of the whole relevant time window. Nevertheless, the refined SPDERG approximation turns out to be the one that makes the largest error, by over-evaluating the correct complex concentration behaviour, in the remaining part of the relevant time window (in detail, for $t \sim 2.5 - 6.5\ s$).

Thus, the present refined SPDERG 2nd order uniform approximations successfully capture the correct dynamical behaviours in a large part of the relevant time windows, despite that the considered case ($\varepsilon \sim 0.5$) is very unfavourable. First of all, this further confirms both the correctness and the utility of the SPDERG approach in general. Moreover, these findings support the effectiveness of the present proposed method for obtaining asymptotically vanishing solutions, that exploits as much as possible the results, too. At least within our way of applying the SPDERG, the proposed refined uniform approximations appear easily generalizable to other similar cases. At the same time, the present analysis makes clear that the remaining part, to be analytically calculated, of the 2nd order outer contributions, would be important for an approximation to MM kinetics, beyond the sQSSA, that could give correct results in the whole relevant time window, also in cases as unfavourable as the here considered one, from the point of view of the values of the kinetic constants and of the expansion parameter.

3.6 The tQSSA framework

To apply the SPDERG approach within the tQSSA framework, is left for future work. Nonetheless, we recall this framework in some detail [29, 81], since it represents the basis for the planned SPDERG application. In particular, this allows us to clarify why such a framework is more interesting from

the experimental point of view and to outline possible simplifications and possible difficulties in the SPDERG application.

In fact, it is only from the Sixties of the last century, that mathematicians have interpreted the sQSSA in terms of leading order in asymptotic expansions with respect to an appropriate parameter, ε, that must be supposed sufficiently small. In the present work, we used $\varepsilon = e_0/s_0$, following [31]. In [92] it was considered $\varepsilon = e_0/(s_0 + K_M)$, showing that the sQSSA is valid in a more extended parameter range than the one supposed by biochemists.

The work by Laidler [82] and the subsequent ones by various other authors [83, 84, 98, 79, 80, 78] (see also [29], and reference therein), introduced a different approach to the study of enzyme kinetics, that is known as total QSSA (tQSSA) and that is valid in a wider range of parameters.

Here we follow the recent work in [81], in which the tQSSA turns out to correspond to the leading order in the expansion parameter $\varepsilon = Ke_0/(K_M + e_0 + s_0)^2$. Importantly, in this work, the first order terms of the inner and outer solutions are computed, and the corresponding uniform approximations are given. This gives us in principle the possibility to check the correctness of the SPDERG results up to the first order in the standard perturbation expansion.

In order to understand the main difference with the previously recalled sQSSA, the starting point is that in the tQSSA one first of all considers as independent variables the substrate and the complex. Hence, one faces the system of two differential equations:

$$\left\{ \begin{array}{lcl} \dot{s}(t) & = & -k_1(e_0 - c(t))s(t) + k_{-1}c(t) \\ \dot{c}(t) & = & k_1(e_0 - c(t))s(t) - (k_{-1} - k_2)c(t), \end{array} \right. \tag{3.77}$$

with initial conditions $s(0) = s_0$, $c(0) = 0$, and conservation laws $e(t)+c(t) = e_0$, $s(t)+c(t)+p(t) = s_0$.

The subsequent key passage consists in the introduction of the *total* substrate $\overline{s}(t) = s(t) + c(t)$. Correspondingly, one studies the system:

$$\left\{ \begin{array}{lcl} \dot{\overline{s}}(t) & = & -k_2 c(t) \\ \dot{c}(t) & = & k_1 \left[c^2(t) - \overline{s}(t)c(t) - (e_0 + K_M)c(t) + e_o\overline{s}(t) \right], \end{array} \right. \tag{3.78}$$

with initial conditions $\overline{s}(0) = s_0$, $c(0) = 0$, and conservation laws $e(t) + c(t) = e_0$, $\overline{s}(t) + p(t) = s_0$.

One can show [81] that the inner equations, for the dimensionless total substrate and complex concentrations, are given by:

$$\left\{ \begin{array}{lcl} \dot{\tilde{\overline{s}}}(\tau) & = & -\varepsilon\tilde{c}(\tau) \\ \dot{\tilde{c}}(\tau) & = & k_1 \left[\sigma\eta\tilde{c}^2(\tau) - \sigma\tilde{\overline{s}}(\tau)\tilde{c}(\tau) - (\eta + \mathcal{K}_M)\tilde{c}(\tau) + \tilde{\overline{s}}(\tau) \right], \end{array} \right. \tag{3.79}$$

with initial conditions $\tilde{\overline{s}}(0) = 1$, $\tilde{c}(0) = 0$.

Here, the introduced parameters are:

$$\sigma = \frac{s_0}{K_M + e_0 + s_0}; \qquad \eta = \frac{e_0}{K_M + e_0 + s_0}; \qquad \mathcal{K}_M = \frac{K_M}{K_M + e_0 + s_0}; \tag{3.80}$$

whereas, as anticipated, $\varepsilon = Ke_0/(K_M + e_0 + s_0)^2$ and one finds $\sigma + \eta + \mathcal{K}_M = 1$. Importantly, the introduced expansion parameter, that arises as the natural perturbation parameter, satisfies $\varepsilon \leq 1/4$ for any set of kinetic constants and initial conditions [29, 81, 99]. This is the reason why the tQSSA, though being more analytically demanding, is a more interesting starting point than the sQSSA from the experimental point of view, particularly in some experimental conditions [29].

The system (3.79) represents the starting point for applying the SPDERG within the tQSSA framework. As we verified in detail in the present work that holds in the case of the SPDERG beyond the sQSSA, one expects in particular that:

- The only quantity to be renormalized is the initial condition on the (total) substrate.

- At each order in the SPDERG approach one finds first of all the leading order terms of the outer solutions in the Perturbation Method.
- In the refined SPDERG uniform approximations, that tend correctly to zero at large times, one finds a part of the terms of the outer solutions at the considered order in the Perturbation Method, too.
- At a given order, the refined SPDERG uniform approximations are better than the previous order perturbation expansion ones, in a region that encompasses the matching one.
- To the aim of calculating the outer contributions one has in principle to solve less demanding ODEs than in the Perturbation Method, particularly at high orders in the expansion parameter.

Finally, notice that, also in the case of the tQSSA framework, the Perturbation Expansion uniform approximations are known only up to the first order in ε. Nonetheless, they have already proved to be very close to the numerical solutions also in demanding cases [29, 81]. Hence, by being able to calculate the SPDERG uniform approximations up to the second order in ε one should definitely get interesting results from the experimental point of view.

The evident difficulty in reaching such an aim is that the Perturbation Method outer contributions are not known explicitly and that they appear analytically involved [81]. In detail, already at the zeroth order, they appear to be non writable in terms of a function as simple as the Lambert one, at least not for any ε. On the other hand, one can think to generalizations of the Lambert function, whose solutions are expected anyway to be easily found numerically.

Chapter 4

Conclusions

It is not easy to conclude the present thesis work since it rather would be the starting point for further studies. Anyway, we obtained various different conclusive results:

- The finite size scaling analyses of data on the considered disordered PS model with $c_p = 2.15$, in terms of pseudo-critical temperatures, definitely confirm that we are extrapolating the correct thermodynamic limit ($\mathcal{N} \to \infty$) behaviour, in which the model undergoes a smooth transition.
  - The mean value of the pseudo-$\mathcal{T}_c$ itself and its fluctuations turn out to be the variables less affected by the strong finite size corrections that characterize such a model, and their study allowed us in particular to get our best estimation of the random fixed point critical exponent, $c_r = 1.35 \pm 0.05$.
  - The study in terms of pseudo-$\mathcal{T}_c$ of the other main observables, in particular of the maximum of the averaged susceptibility, and the careful analyses of data on the probability distribution of the loop length at different temperatures outline the opportunity to introduce a crossover chain length, $\mathcal{N}^*$, in order to describe the finite size model behaviour. We notice that such a point is particularly relevant to DNA denaturation, since one often experimentally observes relatively small molecules.
  - Within a phenomenological scenario, that is based on the importance of the rare regions in the sequence for the smoothening of the transition and that we describe in detail, the knowledge of $\mathcal{N}^*$ for the considered PS model with $c_p = 2.15$ is expected to allow its prediction as a function of the model parameters (in particular of the introduced disorder strength, x).
  - To this aim we rebuild the combinatorial calculation of $P(L,\mathcal{N})$, that is defined just as the probability of obtaining at least L consecutive heads by tossing $\mathcal{N}$ times a fair coin. By definition this is the same probability as the one of obtaining a rare region of minimal length L in a sequence of length $\mathcal{N}$ for the considered uniform binomial distribution of the link energies. Such a calculation allows us to get a refined evaluation of the minimal rare region length, L^*, in the considered model and it represents the basis for the future evaluation of $\mathcal{N}^*(x)$.
- We introduce an *ad hoc* way of applying a Renormalization Group based approach (the SPDERG) to the particularly demanding case of Michaelis-Menten kinetics, by starting from the standard QSSA framework.
  - The SPDERG is proved to be successful in reproducing the correct known terms of the standard Perturbation Method.

– By performing the calculations up to the second order in the expansion variable, ε, we obtain in particular the 2nd order terms of the inner solutions, for the first time to our knowledge.

– By hypothesizing that the observed behaviour of the solution could be iterated, we are able to obtain SPDERG uniform approximations that, besides better reproducing the behaviour of the numerical solutions, in a region encompassing the matching one, than the standard perturbation expansion uniform approximations, tend correctly to zero at large times.

– The study allowed us to outline similarities and differences between the SPDERG and the standard perturbation expansion, both from the analytical and from the theoretical point of view. Importantly, the required imposition of the independence of the solutions on the arbitrary time λ appears equivalent, on the one hand, to apply the scaling condition in the standard Renormalization Group (with λ playing the role of a large time cutoff) and, on the other hand, to impose the matching condition at an "unknown" time in the standard Perturbation Expansion.

– Various outcomes of the present study, such as the detailed verification that the substrate initial condition is the only quantity that needs to be renormalized, can be taken for granted in similar future works.

– In the first place, the present *ad hoc* way of applying the approach should make possible a similar study within the total QSSA framework, that would be particularly interesting from the experimental point of view.

Appendix A

Appendix

A.1 Details of the SIMEX implementation

The SIMEX algorithm, in the present work, is based on the approximation for $1/l^{c_p}$ given by (2.20), *i.e.* $1/(2l)^{2.15} \simeq \sum_{k=1}^{N_E} a_k\, e^{-2lb_k}$, with $N_E = 15$ exponential terms, whereas the values of the coefficients $\{a_k, b_k\}$ (the same as in [38, 39]) are reported in [Tab. 4.1].

The computation of the recursive equations for the forward and backward partition functions within the SIMEX algorithm, here, were implemented with the introduction of free-energy-like quantities, in order to handle logarithms of Z^f and Z^b, by following [66].

In detail [37], the formula (2.14) for the forward partition function becomes:

$$Z^f(\{\epsilon_i\}, n+1, \mathcal{T}) = e^{(\beta\epsilon_{n+1} - \log\mu)} \left\{ Z^f(\{\epsilon_i\}, n, \mathcal{T}) + \sum_{k=1}^{N_E} \sum_{n'=1}^{n-1} Z^f(\{\epsilon_i\}, n', \mathcal{T}) e^{[-2(n-n'+1)b_k]} \right\}. \quad \text{(A.1)}$$

By defining:

$$Q_k(\{\epsilon_i\}, n, \mathcal{T}) = \sum_{n'=1}^{n-1} Z^f(\{\epsilon_i\}, n', \mathcal{T}) e^{2n'b_k} = e^{[2nb_k + \Phi_k^f(\{\epsilon_i\}, n, \mathcal{T})]}, \quad \text{(A.2)}$$

one obtains:

$$Z^f(\{\epsilon_i\}, n, \mathcal{T}) = e^{\Phi_k^f(\{\epsilon_i\}, n, \mathcal{T})} - e^{[-2nb_k + \Phi_k^f(\{\epsilon_i\}, n-1, \mathcal{T})]}, \quad \text{(A.3)}$$

and a recursion relation for the $\Phi_k^f(\{\epsilon_i\}, n, \mathcal{T})$, that are the (forward) free-energy-like quantities:

$$\begin{aligned} \Phi_k^f(\{\epsilon_i\}, n+1, \mathcal{T}) &= \Phi_k^f(\{\epsilon_i\}, n, \mathcal{T}) + \log[A^f(\mathcal{T}) + B^f(\{\epsilon_i\}, n, n-1, \mathcal{T}) + \\ &+ C^f(\{\epsilon_i\}, n, n-1, \mathcal{T})], \end{aligned} \quad \text{(A.4)}$$

with:

$$A^f(\mathcal{T}) = e^{-2b_k} \quad \text{(A.5)}$$

$$B^f(\{\epsilon_i\}, n, n-1, \mathcal{T}) = e^{(\beta\epsilon_{n+1} - \log\mu)} \left\{ 1 - e^{[-2b_k + \Phi_k^f(\{\epsilon_i\}, n-1, \mathcal{T}) - \Phi_k^f(\{\epsilon_i\}, n, \mathcal{T})]} \right\} \quad \text{(A.6)}$$

$$C^f(\{\epsilon_i\}, n, n-1, \mathcal{T}) = e^{[\beta\epsilon_{n+1} - \log\mu - \Phi_k^f(\{\epsilon_i\}, n, \mathcal{T})]} \left\{ \sum_{j=1}^{N_E} a_j e^{[-4b_j + \Phi_j^f(\{\epsilon_i\}, n-1, \mathcal{T})]} \right\}. \quad \text{(A.7)}$$

Analogously, one writes the equation for the backward partition function:

$$Z^b(\{\epsilon_i\}, n-1, \mathcal{T}) = e^{(\beta\epsilon_{n-1} - \log\mu)} \Big\{ Z^b(\{\epsilon_i\}, n, \mathcal{T}) +$$

$$+ \quad \sum_{k=1}^{N_E} \sum_{n'=n+1}^{\mathcal{N}} Z^b(\{\epsilon_i\}, n', \mathcal{T}) e^{[-2(n'-n+1)b_k]} + 1 \Bigg\}, \tag{A.8}$$

and defines:

$$R_k(\{\epsilon_i\}, n, \mathcal{T}) = \sum_{n'=n}^{\mathcal{N}} Z^b(\{\epsilon_i\}, n', \mathcal{T}) e^{-2n'b_k} = e^{[-2nb_k + \Phi_k^b(\{\epsilon_i\}, n, \mathcal{T})]}, \tag{A.9}$$

by obtaining:

$$\begin{aligned} \Phi_k^b(\{\epsilon_i\}, n-1, \mathcal{T}) &= \Phi_k^b(\{\epsilon_i\}, n, \mathcal{T}) + \log[A^b(\mathcal{T}) + B^b(\{\epsilon_i\}, n, n+1, \mathcal{T}) + \\ &+ C^b(\{\epsilon_i\}, n, n+1, \mathcal{T}) + D^b(\{\epsilon_i\}, n, n+1, \mathcal{T})], \end{aligned} \tag{A.10}$$

with:

$$A^b(\mathcal{T}) = e^{-2b_k} \tag{A.11}$$

$$B^b(\{\epsilon_i\}, n, n+1, \mathcal{T}) = e^{(\beta\epsilon_{n-1} - \log\mu)} \left\{ 1 - e^{[-2b_k + \Phi_k^b(\{\epsilon_i\}, n+1, \mathcal{T}) - \Phi_k^b(\{\epsilon_i\}, n, \mathcal{T})]} \right\} \tag{A.12}$$

$$C^b(\{\epsilon_i\}, n, n-1, \mathcal{T}) = e^{[\beta\epsilon_{n-1} - \log\mu - \Phi_k^b(\{\epsilon_i\}, n, \mathcal{T})]} \left\{ \sum_{j=1}^{N_E} a_j e^{[-4b_j + \Phi_j^b(\{\epsilon_i\}, n+1, \mathcal{T})]} \right\} \tag{A.13}$$

$$D^b(\{\epsilon_i\}, n, n-1, \mathcal{T}) = e^{[\beta\epsilon_{n-1} - \log\mu - \Phi_k^b(\{\epsilon_i\}, n, \mathcal{T})]}. \tag{A.14}$$

By moreover making the assumption $Z^f(\{\epsilon_i\}, 0, \mathcal{T}) = Z^b(\{\epsilon_i\}, \mathcal{N}+1, \mathcal{T}) = 0$, we used the boundary conditions:

$$Z^f(\{\epsilon_i\}, 1, \mathcal{T}) = e^{(\beta\epsilon_1 - \log\mu)} \tag{A.15}$$

$$Z^f(\{\epsilon_i\}, 2, \mathcal{T}) = e^{(\beta\epsilon_1 - \log\mu)} e^{(\beta\epsilon_2 - \log\mu)} \tag{A.16}$$

$$Z^b(\{\epsilon_i\}, \mathcal{N}-1, \mathcal{T}) = e^{(\beta\epsilon_{\mathcal{N}} - \log\mu)} \left[e^{(\beta\epsilon_{\mathcal{N}-1} - \log\mu)} + 1 \right] \tag{A.17}$$

$$Z^b(\{\epsilon_i\}, \mathcal{N}, \mathcal{T}) = e^{(\beta\epsilon_{\mathcal{N}} - \log\mu)}. \tag{A.18}$$

Therefore, we get correspondingly:

$$\Phi_k^f(\{\epsilon_i\}, 1, \mathcal{T}) = \beta\epsilon_1 - \log\mu \tag{A.19}$$

$$\Phi_k^f(\{\epsilon_i\}, 2, \mathcal{T}) = \beta\epsilon_1 - \log\mu + \log\left[e^{-2b_k} + e^{(\beta\epsilon_2 - \log\mu)} \right] \tag{A.20}$$

$$\Phi_k^b(\{\epsilon_i\}, \mathcal{N}-1, \mathcal{T}) = \beta\epsilon_{\mathcal{N}} - \log\mu + \log\left\{ e^{-2b_k} + e^{(\beta\epsilon_{\mathcal{N}-1} - \log\mu)} \left[1 + e^{(-\beta\epsilon_{\mathcal{N}} + \log\mu)} \right] \right\} \tag{A.21}$$

$$\Phi_k^b(\{\epsilon_i\}, \mathcal{N}, \mathcal{T}) = \beta\epsilon_{\mathcal{N}} - \log\mu. \tag{A.22}$$

k	a_k	b_k
1	7.4210949671864474	2.469385023938319
2	0.7363845614102907	0.9006571079384265
3	0.08732482111472176	0.33060668412882793
4	$9.887211230992947 \cdot 10^{-3}$	0.1196018051115684
5	$1.102823738631329 \cdot 10^{-3}$	0.0430783130904746
6	$1.2250560616947666 \cdot 10^{-4}$	0.015497408698374987
7	$1.3593126590298345 \cdot 10^{-5}$	$5.573286954347416 \cdot 10^{-3}$
8	$1.5077329395704116 \cdot 10^{-6}$	$2.0039933412704532 \cdot 10^{-3}$
9	$1.6718473281512066 \cdot 10^{-7}$	$7.204015696886597 \cdot 10^{-4}$
10	$1.852449196083061 \cdot 10^{-8}$	$2.5878821693784586 \cdot 10^{-4}$
11	$2.0478548398679403 \cdot 10^{-9}$	$9.275640096786728 \cdot 10^{-5}$
12	$2.247364959164715 \cdot 10^{-10}$	$3.301049580630032 \cdot 10^{-5}$
13	$2.4079053140360544 \cdot 10^{-11}$	$1.1480239104095493 \cdot 10^{-5}$
14	$2.369755720744333 \cdot 10^{-12}$	$3.693407020570475 \cdot 10^{-6}$
15	$1.608846567970496 \cdot 10^{-13}$	$8.820504064582594 \cdot 10^{-7}$

Table A.1: Values of the coefficients used in the SIMEX approximation for $1/l^{2.15}$ with $N_E = 15$ exponentials (see the text for details).

Appendix B

Appendix

B.1 The complex 1st order SPDERG inner solution

We report the solution $\tilde{c}_1^{rg}(\tau)$ of the system given by Eqs. (3.44), *i.e.*, the 1st order inner solution for the dimensionless complex concentration in MM kinetics, beyond the sQSSA, within the SPDERG approach. Here, the ICs to be renormalized are fixed to $\tilde{s}^{rg}(\tau_0) = \tilde{s}_0^* + \varepsilon\tilde{s}_1^*$ and $\tilde{c}^{rg}(\tau_0) = \tilde{c}_0^* + \varepsilon\tilde{c}_1^*$, at $\tau = \tau_0$. In detail, we distinguish the terms corresponding to different functions of $(\tau - \tau_0)$. We find:

$$\begin{aligned}
\tilde{c}_1^{rg}(\tau) &= A_{c_1^{rg}}(\tilde{s}_0^*)(\tau-\tau_0) + B_{c_1^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*, \tilde{c}_0^*) + C_{c_1^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*, \tilde{c}_0^*, \tilde{c}_1^*)e^{-(\tilde{s}_0^*+M)(\tau-\tau_0)} + \\
&+ \left[D_{c_1^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*, \tilde{c}_0^*)(\tau-\tau_0) + E_{c_1^{rg}}(\tilde{s}_0^*, \tilde{c}_0^*)(\tau-\tau_0)^2\right] e^{-(\tilde{s}_0^*+M)(\tau-\tau_0)} + \\
&+ F_{c_1^{rg}}(\tilde{s}_0^*, \tilde{c}_0^*)e^{-2(\tilde{s}_0^*+M)(\tau-\tau_0)},
\end{aligned} \tag{B.1}$$

with:

$$\begin{aligned}
A_{c_1^{rg}}(\tilde{s}_0^*) &= -\frac{M(M-m)}{(\tilde{s}_0^*+M)^3}\tilde{s}_0^*; \\
B_{c_1^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*, \tilde{c}_0^*) &= -\frac{M(\tilde{s}_0^*-M+2m)}{(\tilde{s}_0^*+M)^4}\tilde{s}_0^* + \frac{M}{(\tilde{s}_0^*+M)^2}\tilde{s}_1^* + \frac{M(\tilde{s}_0^*+m)}{(\tilde{s}_0^*+M)^3}\tilde{c}_0^*; \\
C_{c_1^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*, \tilde{c}_0^*, \tilde{c}_1^*) &= \tilde{c}_1^* + \frac{M(\tilde{s}_0^*-M+2m)}{(\tilde{s}_0^*+M)^4}\tilde{s}_0^* + \frac{\tilde{s}_0^*+m}{(\tilde{s}_0^*+M)^4}(\tilde{s}_0^*)^2 - \frac{M}{(\tilde{s}_0^*+M)^2}\tilde{s}_1^* + \\
&- \frac{(2\tilde{s}_0^*+M)(\tilde{s}_0^*+m)}{(\tilde{s}_0^*+M)^3}\tilde{c}_0^* + \frac{\tilde{s}_0^*+m}{(\tilde{s}_0^*+M)^2}(\tilde{c}_0^*)^2; \\
D_{c_1^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*, \tilde{c}_0^*) &= -\frac{(\tilde{s}_0^*+m)(\tilde{s}_0^*-M)}{(\tilde{s}_0^*+M)^3}\tilde{s}_0^* + \frac{\tilde{s}_0^*\tilde{s}_1^*}{\tilde{s}_0^*+M} - \frac{\tilde{s}_1^*\tilde{c}_0^*}{(\tilde{s}_0^*+M)^2} + \frac{(\tilde{s}_0^*+m)(2\tilde{s}_0^*-M)}{(\tilde{s}_0^*+M)^2}\tilde{c}_0^* + \\
&- \frac{\tilde{s}_0^*+m}{(\tilde{s}_0^*+M)^2}(\tilde{c}_0^*)^2; \\
E_{c_1^{rg}}(\tilde{s}_0^*, \tilde{c}_0^*) &= -\frac{(M-m)}{2\,(\tilde{s}_0^*+M)^2}(\tilde{s}_0^*)^2 + \frac{(M-m)}{2(\tilde{s}_0^*+M)}\tilde{s}_0^*\tilde{c}_0^*; \\
F_{c_1^{rg}}(\tilde{s}_0^*, \tilde{c}_0^*) &= -\frac{\tilde{s}_0^*+m}{(\tilde{s}_0^*+M)^4}(\tilde{s}_0^*)^2 + 2\frac{\tilde{s}_0^*+m}{(\tilde{s}_0^*+M)^3}\tilde{s}_0^*\tilde{c}_0^* - \frac{\tilde{s}_0^*+m}{(\tilde{s}_0^*+M)^2}(\tilde{c}_0^*)^2.
\end{aligned} \tag{B.2}$$

In particular, one can check that it is, correctly, $\tilde{c}_1^{rg}(\tau_0) = B_{c_1^{rg}} + C_{c_1^{rg}} + F_{c_1^{rg}} = \tilde{c}_1^*$. Moreover, for $\tilde{s}_0^* = 1$, $\tilde{s}_1^* = \tilde{c}_0^* = \tilde{c}_1^* = 0$ and $\tau_0 = 0$, the result on $\tilde{c}_1^{in}$ reported in (3.21) is correctly reproduced, too.

B.2 The substrate and complex 2nd order SPDERG inner solutions

We report the solutions $\tilde{s}_2^{rg}(\tau)$ and $\tilde{c}_2^{rg}(\tau)$ of the system given by Eqs. (3.59), *i.e.*, the 2nd order inner solutions for the dimensionless substrate and complex concentrations in MM kinetics, beyond the sQSSA, within the SPDERG approach. Here, the ICs are anyway already fixed, for the sake of simplicity, to $\tilde{s}^{rg}(\tau_0) = \tilde{s}_0^* + \varepsilon \tilde{s}_1^*$ and $\tilde{c}^{rg}(\tau_0) = 0$, at $\tau = \tau_0$. As discussed in the text, this choice is made on the basis of the expectation, to be verified, that the only quantities that need to be renormalized at the present order are $\tilde{s}_0^*$ (in agreement with the previously obtained result) and $\tilde{s}_1^*$.

Let us start with $\tilde{s}_2^{rg}(\tau)$. Here, we separate the terms that are different functions of $(\tau - \tau_0)$ and the terms that depend only on $\tilde{s}_0^*$, or both on $\tilde{s}_0^*$ and on $\tilde{s}_1^*$:

$$\begin{aligned}
\tilde{s}_2^{rg}(\tau) &= \left[A_{s_2^{rg}}(\tilde{s}_0^*) + B_{s_2^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*)\right](\tau - \tau_0) + C_{s_2^{rg}}(\tilde{s}_0^*)(\tau - \tau_0)^2 + D_{s_2^{rg}}(\tilde{s}_0^*) + E_{s_2^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*) + \\
&+ \left\{\left[F_{s_2^{rg}}(\tilde{s}_0^*) + G_{s_2^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*)\right] + \left[H_{s_2^{rg}}(\tilde{s}_0^*) + I_{s_2^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*)\right](\tau - \tau_0)\right\} e^{-(\tilde{s}_0^* + M)(\tau - \tau_0)} + \\
&+ J_{s_2^{rg}}(\tilde{s}_0^*)(\tau - \tau_0)^2 e^{-(\tilde{s}_0^* + M)(\tau - \tau_0)} + K_{s_2^{rg}}(\tilde{s}_0^*) e^{-2(\tilde{s}_0^* + M)(\tau - \tau_0)}. \qquad \text{(B.3)}
\end{aligned}$$

In fact, only the first three terms contribute to the part to be renormalized of the whole function, $\tilde{s}_{div}^{rg}(\tau)$, at the 2nd order. Their coefficients are:

$$\begin{aligned}
A_{s_2^{rg}}(\tilde{s}_0^*) &= \frac{2M(M-m)\tilde{s}_0^*}{(\tilde{s}_0^* + M)^4}(\tilde{s}_0^* + m); \qquad B_{s_2^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*) = -\frac{M(M-m)\tilde{s}_1^*}{(\tilde{s}_0^* + M)^2}; \\
C_{s_2^{rg}}(\tilde{s}_0^*) &= \frac{M(M-m)^2\tilde{s}_0^*}{2(\tilde{s}_0^* + M)^3}. \qquad \text{(B.4)}
\end{aligned}$$

For the sake of completeness, we give the explicit dependence on $\tilde{s}_0^*$ and $\tilde{s}_1^*$ of the other terms, too. One finds:

$$\begin{aligned}
D_{s_2^{rg}}(\tilde{s}_0^*) &= \frac{\tilde{s}_0^*}{2(\tilde{s}_0^* + M)^5}\left[(\tilde{s}_0^*)^2(3M - m) - \tilde{s}_0^*(2M^2 - 5Mm - m^2) - 2Mm(2M - 3m)\right]; \\
E_{s_2^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*) &= -G_{s_2^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*) = -\frac{\tilde{s}_1^*}{(\tilde{s}_0^* + M)^3}\left[\tilde{s}_0^*(2M - m) + Mm\right]; \\
F_{s_2^{rg}}(\tilde{s}_0^*) &= -\frac{\tilde{s}_0^*}{(\tilde{s}_0^* + M)^5}\left[(\tilde{s}_0^*)^3 + (\tilde{s}_0^*)^2(2M + m) - \tilde{s}_0^*(M^2 - 3Mm - m^2) - Mm(2M - 3m)\right]; \\
H_{s_2^{rg}}(\tilde{s}_0^*) &= \frac{\tilde{s}_0^*}{(\tilde{s}_0^* + M)^4}\left[(\tilde{s}_0^*)^3 - (\tilde{s}_0^*)^2(M - 2m) - \tilde{s}_0^* M^2 - Mm^2\right]; \\
I_{s_2^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*) &= -\frac{\tilde{s}_0^*\tilde{s}_1^*}{(\tilde{s}_0^* + M)^2}(\tilde{s}_0^* + m); \qquad J_{s_2^{rg}}(\tilde{s}_0^*) = \frac{(\tilde{s}_0^*)^2(M - m)}{2(\tilde{s}_0^* + M)^3}(\tilde{s}_0^* + m); \\
K_{s_2^{rg}}(\tilde{s}_0^*) &= \frac{(\tilde{s}_0^*)^2(\tilde{s}_0^* + m)}{2(\tilde{s}_0^* + M)^5}(2\tilde{s}_0^* + M + m). \qquad \text{(B.5)}
\end{aligned}$$

One can check that, correctly, $\tilde{s}_2^{rg}(\tau_0) = D_{s_2^{rg}} + E_{s_2^{rg}} + F_{s_2^{rg}} + G_{s_2^{rg}} + K_{s_2^{rg}} = 0$. Furthermore, when calculating the $\mathcal{R}_2^s(\tau)$ contribution to the SPDERG 2nd order uniform approximation, with the renormalized divergent part, one is interested in evaluating this quantity in $\tau_0 = 0$, for $\tilde{s}_0^* = 1$ and $\tilde{s}_1^* = 0$. Hence, one immediately gets $E_{s_2^{rg}} = G_{s_2^{rg}} = I_{s_2^{rg}} = 0$, since all of these coefficients are proportional to $\tilde{s}_1^*$. The other coefficients can be easily calculated for $\tilde{s}_0^* = 1$ and the corresponding $\mathcal{R}_2^s$ is reported in (3.70).

In the case of $\tilde{c}_2^{rg}(\tau)$, since the complete formula is cumbersome, we limit ourselves to report explicitly the dependence on $\tilde{s}_0^*$ and $\tilde{s}_1^*$ only for the part of the whole function that needs to be renormalized (*i.e.*, in the coefficients of the terms that are proportional to $(\tau - \tau_0)$ and to $(\tau - \tau_0)^2$).

Correspondingly, we write from the beginning:

$$\tilde{c}_2^{rg}(\tau) = \tilde{c}_{2,div}^{rg}(\tau) + \mathcal{R}_2^c(\tau). \tag{B.6}$$

In fact, we are collecting in $\tilde{c}_{2,div}^{rg}(\tau)$ just the secular terms:

$$\tilde{c}_{2,div}^{rg}(\tau) = \left[A_{c_{2,div}^{rg}}(\tilde{s}_0^*) + B_{c_{2,div}^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*)\right](\tau - \tau_0) + C_{c_{2,div}^{rg}}(\tilde{s}_0^*)(\tau - \tau_0)^2, \tag{B.7}$$

with, in detail:

$$\begin{aligned} A_{c_{2,div}^{rg}}(\tilde{s}_0^*) &= -\frac{M(M-m)\tilde{s}_0^*}{(\tilde{s}_0^* + M)^6}\left[2(\tilde{s}_0^*)^2 - 5\tilde{s}_0^*(M-m) + M^2 - 3Mm\right]; \\ B_{c_{2,div}^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*) &= \frac{M(M-m)\tilde{s}_1^*}{(\tilde{s}_0^* + M)^4}(2\tilde{s}_0^* - M); \qquad C_{c_{2,div}^{rg}}(\tilde{s}_0^*) = -\frac{M(M-m)^2\tilde{s}_0^*}{2(\tilde{s}_0^* + M)^5}(2\tilde{s}_0^* - M). \end{aligned} \tag{B.8}$$

On the other hand, $\mathcal{R}_2^c(\tau)$ contains all the terms that remain constant or tend to zero in the large $(\tau - \tau_0)$ limit and it gives the 2nd order contribution of the inner solution to the SPDERG uniform approximation (that is $O(\varepsilon^2)$). For $\tilde{s}_0^* = 1$ and $\tilde{s}_1^* = 0$, by also coherently evaluating it in $\tau_0 = 0$, one finds:

$$\begin{aligned} \mathcal{R}_2^c(\tau)\Big|_{\substack{\tilde{s}_0^*=1 \\ \tilde{s}_1^*=0}} &= A_{\mathcal{R}_2^c} + \left[B_{\mathcal{R}_2^c} + C_{\mathcal{R}_2^c}\tau + D_{\mathcal{R}_2^c}\tau^2 + E_{\mathcal{R}_2^c}\tau^3 + F_{\mathcal{R}_2^c}\tau^4\right]e^{-(1+M)\tau} + \\ &+ \left[G_{\mathcal{R}_2^c} + H_{\mathcal{R}_2^c}\tau + I_{\mathcal{R}_2^c}\tau^2\right]e^{-2(1+M)\tau} + J_{\mathcal{R}_2^c}e^{-3(1+M)\tau}. \end{aligned} \tag{B.9}$$

Actually, it is useful to report also the complete dependence on $\tilde{s}_0^*$, $\tilde{s}_1^*$, M and m, of the coefficient $A_{\mathcal{R}_2^c}$ (the constant term in the original $\tilde{c}_2^{rg}(\tau)$), since it allows to propose a refined SPDERG 2nd order uniform approximation. We find:

$$\begin{aligned} A_{\mathcal{R}_2^c}(\tilde{s}_0^*, \tilde{s}_1^*) &= -\frac{M(\tilde{s}_0^*)^4}{(\tilde{s}_0^* + M)^7} + \frac{M(\tilde{s}_0^*)^3(9M - 11m)}{2(\tilde{s}_0^* + M)^7} - \frac{M(\tilde{s}_0^*)^2}{2(\tilde{s}_0^* + M)^7}(12M^2 - 27Mm + 13m^2) + \\ &+ \frac{M^2\tilde{s}_0^*(M^2 - 6mM + 6m^2)}{(\tilde{s}_0^* + M)^7} + \frac{M\tilde{s}_1^*}{(\tilde{s}_0^* + M)^5}[2(\tilde{s}_0^*)^2 - \tilde{s}_0^*(5M - 6m) + M(1 - 2m)] + \\ &- \frac{M(\tilde{s}_1^*)^2}{(\tilde{s}_0^* + M)^3}, \end{aligned} \tag{B.10}$$

whereas the dependence on M and m of all the coefficients, calculated in $\tilde{s}_0^* = 1$ and $\tilde{s}_1^* = 0$, is given by:

$$\begin{aligned} A_{\mathcal{R}_2^c} &= \frac{M^4 - 6M^3(m+1)}{(1+M)^7} + \frac{3M^2(4m^2 + 9m + 3)}{2(1+M)^7} - \frac{M(13m^2 + 11m + 2)}{2(1+M)^7}; \\ B_{\mathcal{R}_2^c} &= -\frac{M^4 - 6M^3(m+1)}{(1+M)^7} - \frac{M^2(6m^2 + 10m + 3)}{(1+M)^7} + \frac{M(m-7) - 9m^2 - m}{4(1+M)^7}; \\ C_{\mathcal{R}_2^c} &= \frac{M^3(2m+1)}{(1+M)^6} - \frac{M^2(3m^2 + 7m + 5)}{(1+M)^6} + \frac{M(12m^2 + 15m + 5)}{2(1+M)^6} + \frac{3m^2 + 3m + 2}{2(1+M)^6}; \\ D_{\mathcal{R}_2^c} &= -\frac{1}{2(1+M)^5}[2M^3 + M^2(m^2 - 6m - 3) + M(4m^2 - m - 3) + 2m^2 + 3m + 1]; \\ E_{\mathcal{R}_2^c} &= \frac{(M-m)}{6(1+M)^4}[M^2 + M(2m+3) - 3(m+1)]; \qquad F_{\mathcal{R}_2^c} = -\frac{(M-m)^2}{8(1+M)^3}; \\ G_{\mathcal{R}_2^c} &= -\frac{M^2(7m+3)}{2(1+M)^7} + \frac{M(13m^2 + 11m + 6)}{2(1+M)^7} + \frac{3m^2 + 2m + 1}{(1+M)^7}; \end{aligned}$$

$$H_{\mathcal{R}_2^c} = \frac{1}{(1+M)^6}[M^2+M(2m^2+m+1)-3m-2];$$
$$I_{\mathcal{R}_2^c} = -\frac{(M-m)}{(1+M)^5}(m+1); \qquad J_{\mathcal{R}_2^c} = -\frac{(m+1)}{4(1+M)^7}(M+3m+4). \tag{B.11}$$

One can check that, correctly, $\mathcal{R}_2^c(0) = A_{\mathcal{R}_2^c} + B_{\mathcal{R}_2^c} + G_{\mathcal{R}_2^c} + J_{\mathcal{R}_2^c} = 0$.

B.3 The study of the complex at the 2nd order within the SPDERG approach

Let us relabel $T_1(\tilde{s}_0^*)$, $T_2(\tilde{s}_0^*)$, $T_3(\tilde{s}_0^*, \tilde{s}_1^*)$ and $T_4(\tilde{s}_0^*)$, the coefficients of the first four terms in $\tilde{c}_{div}^{rg}(\tau)$, at the 2nd order, given by (3.72). Hence, one has:

$$\begin{aligned}\tilde{c}_{div}^{rg}(\tau) &= T_1(\tilde{s}_0^*) + \varepsilon T_2(\tilde{s}_0^*) + \varepsilon T_3(\tilde{s}_0^*, \tilde{s}_1^*) + \varepsilon T_4(\tilde{s}_0^*)(\tau-\tau_0) + \\ &+ \varepsilon^2 A_{c_{2,div}^{rg}}(\tilde{s}_0^*)(\tau-\tau_0) + \varepsilon^2 B_{c_{2,div}^{rg}}(\tilde{s}_0^*, \tilde{s}_1^*)(\tau-\tau_0) + \varepsilon^2 C_{c_{2,div}^{rg}}(\tilde{s}_0^*)(\tau-\tau_0)^2,\end{aligned} \tag{B.12}$$

with:

$$\begin{aligned}T_1(\tilde{s}_0^*) &= \frac{\tilde{s}_0^*}{\tilde{s}_0^*+M}; \qquad T_2(\tilde{s}_0^*) = -\frac{M(\tilde{s}_0^*-M+2m)}{(\tilde{s}_0^*+M)^4}\tilde{s}_0^*; \\ T_3(\tilde{s}_0^*, \tilde{s}_1^*) &= \frac{M\tilde{s}_1^*}{(\tilde{s}_0^*+M)^2}; \qquad T_4(\tilde{s}_0^*) = -\frac{M(M-m)\tilde{s}_0^*}{(\tilde{s}_0^*+M)^3};\end{aligned} \tag{B.13}$$

Here, in detail: T_1 is the coefficient of the 0th order term, that was already present in $\tilde{c}_0^{rg}(\tau)$, given by (3.43); $T_2 + T_3 = B_{c_1^{rg}}$, with $B_{c_1^{rg}}$ (calculated in $\tilde{c}_0^* = 0$) the constant term in $\tilde{c}_1^{rg}(\tau)$, given by (B.2); $T_4 = A_{c_1^{rg}}$, with $A_{c_1^{rg}}$ the coefficient of the single 1st order secular term in $\tilde{c}_1^{rg}(\tau)$, given again by (B.2); $A_{c_{2,div}^{rg}}$, $B_{c_{2,div}^{rg}}$ and $C_{c_{2,div}^{rg}}$, are instead the coefficients of the three 2nd order secular terms in $\tilde{c}_2^{rg}(\tau)$, given by (B.8), respectively.

When renormalizing the bare substrate IC, $\tilde{s}^* = \tilde{s}_0^* + \tilde{s}_1^*$, by $\tilde{s}_0^* = (1 + \varepsilon z_{s_0,1} + \varepsilon^2 z_{s_0,2})\tilde{s}_0^{rg*}$ and $\tilde{s}_1^* = (1 + \varepsilon z_{s_0,1})\tilde{s}_1^{rg*}$, respectively (with the already chosen $z_{s_0,1}$, given by (3.48), and $z_{s_0,2}$, $z_{s_1,1}$, given by (3.65)), one finds (up to order ε^2):

$$\begin{aligned}T_1(\tilde{s}_0^*) - T_1(\tilde{s}_0^{rg*}) &= \left[\frac{dT_1(\tilde{s}_0^*)}{d\tilde{s}_0^*}\bigg|_{\tilde{s}_0^*=\tilde{s}_0^{rg*}}\right](\varepsilon z_{s_0,1} + \varepsilon^2 z_{s_0,2})\tilde{s}_0^{rg*} + \frac{1}{2}\left[\frac{d^2T_1(\tilde{s}_0^*)}{d(\tilde{s}_0^*)^2}\bigg|_{\tilde{s}_0^*=\tilde{s}_0^{rg*}}\right](\varepsilon z_{s_0,1})^2(\tilde{s}_0^{rg*})^2 = \\ &= \varepsilon\frac{M(M-m)\tilde{s}_0^{rg*}}{\left(\tilde{s}_0^{rg*}+M\right)^3}(\lambda-\tau_0) - \varepsilon^2 Mm\frac{(M-m)\tilde{s}_0^{rg*}}{\left(\tilde{s}_0^{rg*}+M\right)^5}(\lambda-\tau_0) + \\ &- \varepsilon^2 M\frac{(M-m)^2\tilde{s}_0^{rg*}}{2\left(\tilde{s}_0^{rg*}+M\right)^5}(2\tilde{s}_0^{rg*}-M)(\lambda-\tau_0)^2 = \\ &= -\varepsilon T_4(\tilde{s}_0^{rg*})(\lambda-\tau_0) - \varepsilon^2 Mm\frac{(M-m)\tilde{s}_0^{rg*}}{\left(\tilde{s}_0^{rg*}+M\right)^5}(\lambda-\tau_0) + \varepsilon^2 C_{c_{2,div}^{rg}}(\tilde{s}_0^{rg*})(\lambda-\tau_0)^2;\end{aligned} \tag{B.14}$$

$$\begin{aligned}\varepsilon\left[T_2(\tilde{s}_0^*) - T_2(\tilde{s}_0^{rg*})\right] &= \varepsilon\left[\frac{dT_2(\tilde{s}_0^*)}{d\tilde{s}_0^*}\bigg|_{\tilde{s}_0^*=\tilde{s}_0^{rg*}}\right](\varepsilon z_{s_0,1})\tilde{s}_0^{rg*} = \\ &= \varepsilon^2\frac{M(M-m)\tilde{s}_0^{rg*}}{\left(\tilde{s}_0^{rg*}+M\right)^6}\left[2(\tilde{s}_0^{rg*})^2 - (5M-6m)\tilde{s}_0^{rg*} + M(M-2m)\right](\lambda-\tau_0) = \\ &= -\varepsilon^2 A_{c_{2,div}^{rg}}(\tilde{s}_0^{rg*})(\lambda-\tau_0) + \varepsilon^2 Mm\frac{(M-m)\tilde{s}_0^{rg*}}{\left(\tilde{s}_0^{rg*}+M\right)^5}(\lambda-\tau_0);\end{aligned} \tag{B.15}$$

$$\varepsilon\big[T_3(\tilde{s}_0^*,\tilde{s}_1^*)-T_3(\tilde{s}_0^{rg*},\tilde{s}_1^{rg*})\big]=\varepsilon\left[\frac{\partial T_3(\tilde{s}_0^*,\tilde{s}_1^*)}{\partial\tilde{s}_0^*}\bigg|_{\substack{\tilde{s}_0^*=\tilde{s}_0^{rg*}\\ \tilde{s}_1^*=\tilde{s}_1^{rg*}}}\right](\varepsilon z_{s_0,1})\tilde{s}_0^{rg*}+\varepsilon\left[\frac{\partial T_3(\tilde{s}_0^*,\tilde{s}_1^*)}{\partial\tilde{s}_1^*}\bigg|_{\substack{\tilde{s}_0^*=\tilde{s}_0^{rg*}\\ \tilde{s}_1^*=\tilde{s}_1^{rg*}}}\right](\varepsilon z_{s_1,1})\tilde{s}_1^{rg*}=$$
$$=\varepsilon^2\left[-\frac{2M(M-m)\tilde{s}_1^{rg*}}{\big(\tilde{s}_0^{rg*}+M\big)^4}\tilde{s}_0^{rg*}+\frac{M^2(M-m)}{\big(\tilde{s}_0^{rg*}+M\big)^4}\tilde{s}_1^{rg*}\right](\lambda-\tau_0)=$$
$$=-\varepsilon^2 B_{c_{2,div}^{rg}}(\tilde{s}_0^{rg*},\tilde{s}_1^{rg*})(\lambda-\tau_0); \tag{B.16}$$

$$\varepsilon\left[T_4(\tilde{s}_0^*)-T_4(\tilde{s}_0^{rg*})\right]=\varepsilon\left[\frac{dT_4(\tilde{s}_0^*)}{d\tilde{s}_0^*}\bigg|_{\tilde{s}_0^*=\tilde{s}_0^{rg*}}\right](\varepsilon z_{s_0,1})\tilde{s}_0^{rg*}=\varepsilon^2\frac{M(M-m)^2\tilde{s}_0^{rg*}}{\big(\tilde{s}_0^{rg*}+M\big)^4}(2\tilde{s}_0^{rg*}-M)(\lambda-\tau_0)=$$
$$=-2\varepsilon^2 C_{c_{2,div}^{rg}}(\tilde{s}_0^{rg*})(\lambda-\tau_0). \tag{B.17}$$

Therefore, when moreover writing $(\tau-\tau_0)=(\tau-\lambda)+(\lambda-\tau_0)$, since obviously $(\tau-\tau_0)^2=(\tau-\lambda)^2+2(\tau-\lambda)(\lambda-\tau_0)+(\lambda-\tau_0)^2$, one gets exactly the same form of $\tilde{c}_{div}^{rg}(\tau,\lambda)$ given by (3.72), with $\tau_0\to\lambda$, $\tilde{s}_0^*\to\tilde{s}_0^{rg*}(\lambda)$ and $\tilde{s}_1^*\to\tilde{s}_1^{rg*}(\lambda)$.

Let us now remind that the 1st order scaling condition on the renormalized substrate IC, given by Eqs. (3.51), implies the relation $d\tilde{s}_0^{rg*}/d\lambda=-\varepsilon z_{s_0,1}\tilde{s}_0^{rg*}/(\lambda-\tau_0)$, too. Correspondingly, one can make partially use again of the previous formulas in the study of the derivative with respect to λ of $\tilde{c}_{div}^{rg}(\tau,\lambda)$:

$$\begin{aligned}\frac{d\tilde{c}_{div}^{rg}(\tau,\lambda)}{d\lambda} &= \frac{dT_1(\tilde{s}_0^{rg*})}{d\tilde{s}_0^{rg*}}\frac{d\tilde{s}_0^{rg*}}{d\lambda}+\varepsilon\frac{dT_2(\tilde{s}_0^{rg*})}{d\tilde{s}_0^{rg*}}\frac{d\tilde{s}_0^{rg*}}{d\lambda}+\varepsilon\frac{\partial T_3(\tilde{s}_0^{rg*},\tilde{s}_1^{rg*})}{\partial\tilde{s}_0^{rg*}}\frac{d\tilde{s}_0^{rg*}}{d\lambda}+\\
&+\varepsilon\frac{\partial T_3(\tilde{s}_0^{rg*},\tilde{s}_1^{rg*})}{\partial\tilde{s}_1^{rg*}}\frac{d\tilde{s}_1^{rg*}}{d\lambda}+\varepsilon\frac{dT_4(\tilde{s}_0^{rg*})}{d\tilde{s}_0^{rg*}}\frac{d\tilde{s}_0^{rg*}}{d\lambda}(\tau-\lambda)-\varepsilon T_4(\tilde{s}_0^{rg*})+\\
&-\varepsilon^2 A_{c_{2,div}^{rg}}(\tilde{s}_0^{rg*})-\varepsilon^2 B_{c_{2,div}^{rg}}(\tilde{s}_0^{rg*},\tilde{s}_1^{rg*})-2\varepsilon^2 C_{c_{2,div}^{rg}}(\tilde{s}_0^{rg*})(\tau-\lambda).\end{aligned} \tag{B.18}$$

First of all, in order to get $d\tilde{c}_{div}^{rg}(\tau,\lambda)/d\lambda=0$, the 1st order known result on $d\tilde{s}_0^{rg*}/d\lambda$ needs to be once again satisfied. From this point of view, see in particular Eqs. (B.14): indeed, the contribution that comes from the term $dT_1(\tilde{s}_0^{rg*})/d\tilde{s}_0^{rg*}$, in the same way as the first term in that formula, and the one that is equal to $-\varepsilon T_4(\tilde{s}_0^{rg*})$ are the only contributions to be proportional to ε; thus, they obviously need to cancel each other.

Then, by further exploiting the just recalled relation, $d\tilde{s}_0^{rg*}/d\lambda=-\varepsilon z_{s_0,1}\tilde{s}_0^{rg*}/(\lambda-\tau_0)$, one finds that the term involving the derivative of T_2 is partially cancelled by the one proportional to $A_{c_{2,div}^{rg}}$, leaving a contribution equal to $-\varepsilon^2 Mm(M-m)\tilde{s}_0^{rg*}/(\tilde{s}_0^{rg*}+M)^5$, as can be seen from Eqs. (B.15). Moreover, the two terms proportional to $(\tau-\lambda)$ cancel each other, too, as can be seen from Eqs. (B.17).

Therefore, by using $d\tilde{s}_0^{rg*}(\lambda)/d\lambda=-\varepsilon(M-m)\tilde{s}_0^{rg*}/(\tilde{s}_0^{rg*}+M)$, by imposing the scaling condition $d\tilde{c}_{div}^{rg}(\tau,\lambda)/d\lambda=0$ at the 2nd order and by making partially use of Eqs. (B.16), we end up with the equation:

$$-\varepsilon\frac{Mm(M-m)\tilde{s}_0^{rg*}}{(\tilde{s}_0^{rg*}+M)^5}+\varepsilon\frac{2M(M-m)\tilde{s}_0^{rg*}\tilde{s}_1^{rg*}}{\big(\tilde{s}_0^{rg*}+M\big)^4}+\frac{M}{(\tilde{s}_0^{rg*}+M)^2}\frac{d\tilde{s}_1^{rg*}}{d\lambda}-\varepsilon\frac{M(M-m)\tilde{s}_1^{rg*}}{(\tilde{s}_0^{rg*}+M)^4}(2\tilde{s}_0^{rg*}-M)=0. \tag{B.19}$$

Hence, we get the expected result:

$$\frac{d\tilde{s}_1^{rg*}}{d\lambda}=\varepsilon\left[\frac{m(M-m)\tilde{s}_0^{rg*}}{(\tilde{s}_0^{rg*}+M)^3}-\frac{M(M-m)\tilde{s}_1^{rg*}}{(\tilde{s}_0^{rg*}+M)^2}\right], \tag{B.20}$$

that completes the present verification. In fact, this is the same ODE, to be obeyed by $\tilde{s}_1^{rg*}(\lambda)$, that we previously obtained in the study of the substrate and that is reported in Eqs. (3.67).

Bibliography

[1] C. Mascia and E. Montefusco, *Un Invito alla Biomatematica*, 2014 (Edizioni La Dotta, Bologna).

[2] C.C. Lin and L.A. Segel, *Mathematics Applied to Deterministic Problems in the Natural Sciences*, 1988 (SIAM, Philadelphia).

[3] J.D. Murray, *Mathematical Biology*, Vol. I, 3rd Edition, 2004 (Springer-Verlag, New York).

[4] L. Edelstein-Keshet, *Mathematical Models in Biology*, 2005 (SIAM, Philadelphia).

[5] D. Poland and H.A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers*, 1970 (Academic, New York).

[6] G. Giacomin, *Random Polymers Models*, 2007 (Imperial College Press, London).

[7] K. Huang, *Statistical Mechanics*, 2nd Edition, 1987 (John Wiley & Sons, Canada).

[8] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, Fourth Edition, Clarendon Press, Oxford (2002).

[9] R.J. Creswick, H.A. Farach and C.P. Poole *Introduction to Renormalization Group Methods in Physics*, Wiley, New York (1992).

[10] C.M. Bender and S.A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, 1999 (Springer-Verlag, New York).

[11] D. Poland and H.A. Scheraga, *Phase transitions in one dimension and the helix-coil transition in polyamino acids*, 1966, J. Chem. Phys. **45**, 1456.

[12] D. Poland and H.A. Scheraga, *Occurrence of a phase transition in nucleic acid models*, 1966, J. Chem. Phys. **45**, 1464.

[13] M.E. Fisher, *Walks, walls, wetting, and melting*, 1984, J. Stat. Phys. **34**, 667.

[14] Y. Kafri, D. Mukamel and L. Peliti, *Why is the DNA Denaturation Transition First Order?*, 2000, Phys. Rev. Lett. **85**, 4988. A. Hanke and R. Metzler, *Comment*, 2003, Phys. Rev. Lett. **90**, 159801; Y. Kafri, D. Mukamel and L. Peliti, *Reply*, 2003, Phys. Rev. Lett. **90** 159802.

[15] Y. Kafri, D. Mukamel and L. Peliti, *Melting and unzipping of DNA*, 2002, Eur. Phys. J. B **27**, 135.

[16] B. Duplantier, *Polymer Network of fixed topology: renormalization, exact critical exponent γ in two dimensions, and $d = 4 - \varepsilon$*, 1986, Phys. Rev. Lett. **57**, 941.

[17] B. Duplantier, *Statistical mechanics of polymer networks of any topology*, 1989, J. Stat. Phys. **54**, 581.

[18] J.T. Chayes, L. Chayes, D.S. Fisher and T. Spencer, *Finite-size scaling and correlation lengths for disordered systems*, 1986, Phys. Rev. Lett. **57**, 2999.

[19] A. Aharony and A.B. Harris, *Absence of Self-Averaging and Universal Fluctuations in Random Systems near Critical Points*, 1996, Phys. Rev. Lett. **77**, 3700.

[20] G. Giacomin and F.L. Toninelli, *Smoothing of depinning transitions for directed polymers with quenched disorder*, 2006, Phys. Rev. Lett. **96**, 060702.

[21] G. Giacomin and F.L. Toninelli, *Smoothing effect of quenched disorder on polymer depinning transitions*, 2006, Commun. Math. Phys. **266**, 1.

[22] H. Kunz and R. Livi, *DNA denaturation and wetting in the presence of disorder*, 2012, Europhys. Lett. **99**, 30001.

[23] Y. Hafri and D. Mukamel, *Griffiths singularities in unbinding of strongly disordered polymers*, 2003, Phys. Rev. Lett. **91**, 055502.

[24] C. Monthus, *Strong Disorder Renewal Approach to DNA denaturation and wetting: typical and large deviation properties of the free energy*, 2017, J. Stat. Mech. 013301.

[25] R.M. Wartell and A.S. Benight, *Thermal denaturation of DNA molecules: A comparison of theory with experiment*, 1985, Phys. Rep. **126**, 67.

[26] L.-Y. Chen, N. Goldenfeld and Y. Oono, *Renormalization group theory for global asymptotic analysis*, 1994, Phys. Rev. Lett. **73**, 1311.

[27] L.-Y. Chen, N. Goldenfeld and Y. Oono, *Renormalization group and singular perturbations: Multiple scales, boundary layers, and reductive perturbation theory*, 1996, Phys. Rev. E **54**, 376.

[28] L. Michaelis and M.L. Menten, *Die Kinetic der Invertinwirkung*, 1913, Biochem. Z. **49**, 333; English review and translation in K.A. Johnson and R.S. Goody, *The original Michaelis constant: translation of the 1913 Michaelis-Menten paper.*, 2011, Biochemistry **50**, 8264.

[29] A.M. Bersani, E. Bersani, G. Dell'Acqua and M.G. Pedersen, *New trends and perspectives in nonlinear intracellular dynamics: one century from Michaelis-Menten paper*, 2015, Continuum Mech. Thermodyn. **27**, 659.

[30] K.A. Connors, *Chemical Kinetics: The Study of Reactions Rates in Solution*, 1990 (VCH Publisher, New York).

[31] F.G. Heineken, H.M. Tsushiya and R. Aris, *On the mathematical status of the pseudo-steady-state hypothesis of biochemical kinetics*, 1967, Math. Biosci. **1**, 95.

[32] B. Coluzzi and E. Yeramian, *Numerical study of DNA denaturation with self-avoidance: pseudo-critical temperatures and finite size behaviour* 2016, J. Stat. Mech. 043212.

[33] E. Yeramian and P. Claverie, *Analysis of multiexponential functions without a hypothesis as to the number of components*, 1987, *Nature* **326**, 169.

[34] P. Claverie, A. Denis and E. Yeramian, *The representation of functions through the combined use of integral transforms and Padé approximants: Padé-Laplace analysis of functions as sums of exponential*, 1989, Comput. Phys. Rep. **9**, 247.

[35] M. Fixman and J.J. Freire, *Theory of DNA melting curves*, 1977, Biopolymers **16**, 2693.

[36] B. Coluzzi, *Numerical study of a disordered model for DNA denaturation transition*, 2006, Phys. Rev. E, **73** 011911.

[37] B. Coluzzi and E. Yeramian, *Numerical evidence for relevance of disorder in a Poland-Scheraga DNA denaturation model with self-avoidance: scaling behavior of average quantities*, 2007, Eur. Phys. J. B **56**, 349.

[38] T. Garel and C. Monthus, *Numerical study of the disordered PolandScheraga model of DNA denaturation*, 2005 J. Stat. Mech., P06004.

[39] C. Monthus and T. Garel, *Distribution of pseudo-critical temperatures and lack of self-averaging in disordered Poland-Scheraga models with different loop exponents*, 2005, Eur. Phys. J. B **48**, 393.

[40] B. Coluzzi, A.M. Bersani and E. Bersani, *An alternative approach to Michaelis-Menten kinetics that is based on the Renormalization Group*, 2018, Math. Biosci. **299**, 28-58.

[41] E. Raccah, *Applicazione delle tecniche del gruppo di rinormalizzazione alle cinetiche enzimatiche*, Tesi di Laurea Specialistica in Fisica, Sapienza University of Rome, Italy (A.A. 2011/2012).

[42] D. Jost and R. Everaers, *A unified Poland-Scheraga model of oligo- and polynucleotide DNA melting: salt effects and predictive power*, 2009, Biophys. J. **96**, 1056.

[43] B.H. Zimm and J.K. Bragg, *Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains*, 1959, J. Chem. Phys., **31**, 526.

[44] M.S. Causo, B. Coluzzi and P. Grassberger *Simple model for DNA denaturation transition*, 2000, Phys. Rev. E, **62**, 3958.

[45] M.E. Fisher, *Effect of excluded volume on phase transitions in biopolymers*, 1966, J. Chem Phys. **45**, 1469.

[46] E. Carlon, E. Orlandini and A.L. Stella, *Roles of stiffness and excluded volume in DNA denaturation*, 2002, Phys. Rev. Lett. **88**, 198101.

[47] R. Blossey and E. Carlon, *Reparametrizing the loop entropy weights: Effect on DNA melting curves*, 2003, Phys. Rev. E **68**, 061911.

[48] M. Mézard, G. Parisi and M.A. Virasoro, 1987, *Spin Glass Theory and Beyond: An introduction to the replica method and its applications* (World Scientific, Singapore).

[49] G. Toulose, *Theory of the frustration effect in spin glasses: I*, 1977, Comm. Phys. **2**, 115.

[50] H. Orland and A. Zee *RNA folding and large N matrix theory*, 2002, Nucl. Phys. B **620**, 456.

[51] T.R. Einert, P. N ager, H. Orland, R.R. Netz, *Impact of loop statistics on the thermodynamics of RNA folding*, 2008, Phys. Rev. Lett. **101**, 048103.

[52] F. David and K.J. Wiese *Field theory of the RNA freezing transition*, 2009, J. Stat. Mech. P10019.

[53] T. Hwa, E. Marinari, K. Sneppen and L.-h. Tang, Localization of denaturation bubbles in random DNA sequences 2003, PNAS, **100**, 4441.

[54] W. Selke, L.N. Shchur and A.L. Talapov, *Monte Carlo Simulations of Dilute Ising Models*, in *Annual Reviews of Computational Physics*, 1994, edited by D. Stauffer (World Scientific, Singapore).

[55] A.B. Harris, *Effect of random defects on the critical behaviour of Ising models*, 1974, J. Phys. C **7**, 1671.

[56] M. Aizenman and J. Wehr, *Rounding of first-order phase transitions in systems with quenched disorder*, 1989, Phys. Rev. Lett. **62**, 2503; *Erratum*, 1990, Phys. Rev. Lett. **64**, 1311.

[57] A.N. Berker, *Critical behavior induced by quenched disorder*, 1993, Physica A **194**, 72.

[58] P. Grassberger, *Pruned-enriched Rosenbluth method: Simulations of θ polymers of chain length up to 1 000 000*, 1997, Phys. Rev. E **56**, 3682.

[59] F. Iglói, Yu-Cheng Lin, H. Rieger, C. Monthus, *Finite-size scaling of pseudocritical point distributions in the random transverse-field Ising chain*, 2007, Phys. Rev. B **76**, 064421.

[60] C. Monthus and T. Garel, *Random wetting transition on the Cayley tree: a disordered first-order transition with two correlation length exponents*, 2009, J. Phys. A: Math. Theor. **42**, 165003.

[61] R. Brout, *Statistical Mechanical Theory of a Random Ferromagnetic System*, 1959, Phys. Rev. **115**, 824.

[62] S. Wiseman and E. Domany, *Finite-Size Scaling and Lack of Self-Averaging in Critical Disordered Systems*, 1998, Phys. Rev. Lett. **81**, 22;

[63] S. Wiseman and E. Domany, *Self-averaging, distribution of pseudocritical temperatures, and finite size scaling in critical disordered systems*, 1998, Phys. Rev. E **58**, 2938.

[64] D.S. Fisher, *Random transverse field Ising spin chains*, 1992 Phys. Rev. Lett. **69**, 534

[65] D.S. Fisher, *Critical behavior of random transverse-field Ising spin chains*, 1995, Phys. Rev. B **51**, 6411.

[66] T. Garel and H. Orland, *Generalized Poland-Scheraga model for DNA hybridization*, 2004, Biopolymers, **75**, 453.

[67] E. Yeramian, *et al.*, *An optimal formulation of the matrix method in statistical mechanics of one-dimensional interacting units: Efficient iterative algorithmic procedures*, 1990, Biopolymers **30**, 481.

[68] E. Yeramian, *Complexity and tractability. Statistical mechanics of helix-coil transitions in circular DNA as a model-problem*, 1994, Europhys. Lett. **25**, 49.

[69] E. Yeramian and E. Debonneuil, *Probabilistic sequence alignments: Realistic models with efficient algorithms*, 2007, Phys. Rev. Lett. **98**, 078101.

[70] J. Matoušek and J. Nešetřil, *Invitation to Discrete Mathematics*, 2nd Edition, 2008, Oxford University Press (Oxford).

[71] D.A. Wolfram *Solving generalized Fibonacci recurrences*, 1998, Fibonacci Quart. **36**, 129.

[72] A.F. Horadam, *A Generalized Fibonacci Sequence*, 1961, Amer. Math. Monthly **68**, 455.

[73] M.E. Waddill, *The Tetranacci Sequence and Its Generalizations*, 1992, Fibonacci Quart. **30**, 9.

[74] N.J.A. Sloane, *The on-line encyclopedia of integer sequences*, `https://oeis.org`.

[75] J.J. Bravo, C.A. Gómez and F. Luca, *Powers of two as sums of two k-Fibonacci numbers*, 2016, Miskolc Math. Notes **17**, 85.

[76] K.G. Wilson, *Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture* Phys. Rev. B **4**, 3174.

[77] G. Angiulli, et al, *Leishmania infantum trypanothione reductase is a promiscuous enzyme carrying an NADPH:O_2 oxidoreductase activity shared by glutathione reductase*, 2015, Biochimica et Biophysica Acta **1850**, 1891.

[78] J. Borghans, R. de Boer and L. Segel, *Extending the quasi-steady state approximation by changing variables*, 1996, Bull. Math. Biol. **58**, 43.

[79] A.R. Tzafriri and E.R. Edelman, *The total quasi-steady-state approximation is valid for reversible enzyme kinetics*, 2004, J. Theor. Biol. **226**, 303.

[80] A.R. Tzafriri and E.R. Edelman, *Quasi-steady-state kinetics at enzyme and substrate concentrations in excess of the Michaelis-Menten constant*, 2007, J. Theor. Biol. **245**, 737.

[81] G. Dell'Acqua and A.M. Bersani, *A perturbation solution of Michaelis-Menten kinetics in a "total" framework*, 2012, J. Math. Chem. **50**, 1136.

[82] K.J. Laidler, *Theory of the transient phase in kinetics, with special reference to enzyme systems*, 1955, Can. J. Chem. **33**, 1614.

[83] P.A.T. Swoboda, *The kinetics of enzyme action*, 1957, Biochimica et Biophysica acta **23**, 70;

[84] P.A.T. Swoboda, *The kinetics of enzyme action II. The terminal phase of the reaction*, 1957, Biochimica et Biophysica acta **25**, 132.

[85] V. Henri, *Recherches sur la loi de l'action de la sucrase*, 1901, C. R. Hebd. Acad. Sci. **133**, 891;

[86] V. Henri, *Über das Gesetz der Wirkung des Invertins*, 1901, Z. Phys. Chem., **39**, 194;

[87] V. Henri, *Thèorie gènèrale de l'action de quelques diastases*, 1902, C. R. Hebd. Acad. Sci. **135**, 916.

[88] G.E. Briggs and J.B.S. Haldane, *A note on the kinetics of enzyme action*, 1925, Biochem. J. **19**, 338.

[89] D.D. Van Slyke and G.E. Cullen, *The mode of action of urease and of enzymes in general*, 1914, J. Biol. Chem. **19**, 141.

[90] S.J. Fraser, *The steady state and equilibrium approximations: A geometrical picture*, 1988, J. Chem. Phys. **88**, 4732.

[91] S.M. Hanson and S. Schnell, *Reactant Stationary Approximation in Enzyme Kinetics*, 2008, J. Phys. Chem. A **112**, 8654.

[92] L.A. Segel and M. Slemrod, *The quasi steady-state assumption: a case study in perturbation*, 1989, SIAM Rev. **31**, 446.

[93] S. Schnell and C. Mendoza, *A closed-form solution for time-dependent enzyme kinetic*, 1997, J. Theor. Biol. **187**, 207.

[94] R.M. Corless, *et al*, *On the Lambert W function*, 1996, Adv. Comput. Math. **5**, 329.

[95] A.B. Vasil'eva, *Asymptotic behaviour of solutions to certain problems involving non-linear differential equations containing a small parameter multiplying the highest derivatives*, 1963, Uspekhi Mat. Nauk. **18**, 15, in Russian; English translation in Russ. Math. Surv. **18**, 13.

[96] F.C. Hoppensteadt, *Singular perturbations on the infinite interval*, Trans. Am. Math. Soc. **123**, 521-535, 1966.

[97] M.F. Morales and D.E. Goldman, *A Note on the Differential Equation of Simple Enzyme Kinetics*, 1955, J. Am. Chem. Soc. **77**, 6069.

[98] M. Schauer and R. Heinrich, *Analysis of the quasi-steady-state approximation for an enzymatic one-sub- strate reaction*, 1979, J. Theor. Biol. **79**, 425.

[99] J.W. Dingee, A.B. Anton, *A new perturbation solution to the Michaelis-Menten problem*, 2008, AIChe J. **54**, 1344.