Chapter 2

Modeling DNA structure: the *twist-opening* model

2.1 Introduction

As discussed in the previous chapter, we are interested in a DNA model which reproduces and predicts the dynamical features involved in some biological processes, namely in transcription and transcription initiation.

During the last twenty years, after the first work of Englander *et al.* [7], many authors[1] have proposed non linear DNA dynamical models, mainly based on the idea that in a DNA chain vibrational energy could be localized in excitations of the solitary wave type. These models introduce generally one or two degrees of freedom per base pair, as *e.g.* hydrogen bond stretching, base pairs distance along the molecular axis, single base rotations around the backbone axis. Most of these models describe the DNA chain in a continuum approximation leading to equations possessing traveling soliton solutions[2]. In some cases, these solitons could be interpreted as local (moving) base pair breaking regions, as requested in the frame of a description of open complex formation. The continuum approximation seems anyway not appropriate in describing an intrinsically discrete object like DNA.

Our reference model will be the Peyrard Bishop (PB) model. In this chapter we first describe this model and its main results, explaining why it could represent a good starting point for our aims (Section 2.2). We point out its deficiencies and which are the necessary improvements in order to obtain a good representation of the double helix dynamics (Section 2.3). We introduce at this point the *twist-opening* model, explaining how it leads to a correct representation of the double helix structure and stressing the coupling properties between its two

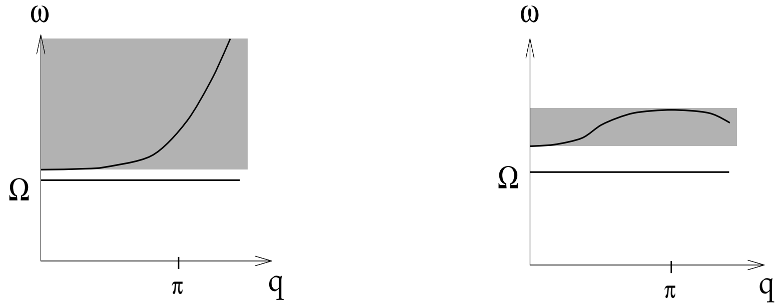
[1] See [13] for a complete review.

[2] The term *soliton* indicates, properly, non linear localized, standing and traveling, exact solutions of completely integrable systems.

degrees of freedom (Section 2.4). We then discuss in detail how to chose the model parameters on the basis of the experimental results (Section 2.5). Finally, we introduce an approximation of the equations of motion for the *twist-opening* model, written in a rescaled form, valid for small displacements from equilibrium (Section 2.6).

A reference discrete model: the PB model

An interesting approach has been followed by Peyrard and Bishop(*PB* in the following) [14], who have proposed a simple DNA planar model where the only degrees of freedom are the hydrogen bond lengths. Its principal aim is to study the dynamics and the thermodynamics of base pairs openings in denaturation and transcription. The PB model will be the starting point for the construction of our model. Besides the great deal of results obtained by Peyrard and co-workers in describing the dynamical behavior and the thermal formation of localized excitations of thermalized DNA chains, one of the important interests of their work is in fact that they pay a special attention in considering discreteness effects.



(a) *(b)*

Figure 2.1: Existence of breathers in discrete systems. Curves represents the linear frequency $\omega(q)$. **(a)**: sketch of the unbounded dispersion curve for a continuous system. **(b)**: sketch of the bounded spectrum of a discrete system. Dark indicates the phonon band.

Discrete lattices can sustain time-periodic, oscillating, space-localized solutions, or *discrete breathers*[3], whose existence is a consequence of the boundedness of linear frequency spectra for discrete systems. Breather stability in discrete lattices arises from the fact that their oscillation frequency Ω lies outside the linear spectrum of the system, so that they cannot exchange energy by resonance with the excitations in the phonon band. Even if it exists a minimum frequency

[3]Also called in the literature intrinsic localized modes, self-localized anharmonic modes or non linear localized excitations. Discrete breathers are widely investigated, mainly by numerical methods. See [31] for an exhaustive report.

and then a possible frequency band of existence for the breather, in the case of nonlinear systems harmonics $n\Omega$ of the breather frequency are excited and must also lie outside the phonon band to avoid coupling to linear plane waves. In the continuum limit for which the upper limit of the phonon band extends to infinity, any frequency Ω will have harmonics in the phonon band and the breather cannot exist[4]. In the case of discrete systems the spectrum does not cover all the frequencies, and thus there are frequency regions or gaps where a breather can live without energy dispersion, as sketched in Figure 2.1.

MacKay and Aubry [32] gave an existence proof for discrete breathers in a large class of discrete systems, as extensions of breather solutions which exist in the limit of zero lattice coupling (*anticontinuum or antiintegrable limit*).

Approximate breather solutions

For a discrete lattice, whose motion is described by ordinary differential equations *(ODE)*, exact analytical solutions exist only in some very special cases. More generally, there are just approximate analytical solutions, that can be obtained by making some appropriate hypotheses on the behavior of the solution in space and time. For example, let us consider a discrete (one dimensional) system characterized by a linear dispersion and submitted to a substrate potential. Such a system is governed by (discrete) *Klein-Gordon* equations of the type

$$\ddot{u}_n - (u_{n+1} + u_{n-1} - 2u_n) + f(u_n) = 0 \,. \tag{2.1}$$

Once we restrict ourselves to look for solutions characterized by an internal fast oscillation modulated by a sufficiently slow varying envelope, it is possible to use a continuum approximation to determine the envelope function, while the internal oscillation is still described in the discrete framework. This leads to reduce the problem to a *Non Linear Schrödinger (NLS)* equation for the envelope. The NLS equation admits, for an appropriate choice of parameters, an *envelope soliton solution*, which is an exact breather, *i.e.*, in this case, a continuous, space-localized, time-periodic function $F(x,t)$ satisfying the conditions

$$F(x,t+T) = F(t)\,, \qquad \lim_{|x|\to\infty} F(x,t) = 0 \quad \text{and} \quad \frac{\partial F}{\partial t} \not\equiv 0\,. \tag{2.2}$$

The complete approximate solution that one finally obtains by superimposing the envelope to the internal oscillation is also a breather.

This kind of solutions can be found for the PB model. They can be interpreted as precursors of opening bubbles, in denaturation as well as in transcription. We will discuss in Section 2.2 what is the biological interest of these breather

[4]The sine-Gordon case represents an exception to this rule, due to special orthogonality relations between plane waves and breather high harmonics.

oscillations. This section will be devoted to a summary of the main results obtained by the PB model.

2.2 A planar model

2.2.1 The Peyrard Bishop model

Model purposes

The Peyrard Bishop model [14] is an attempt to describe physically the dynamics of the formation of transcription bubbles. Being the starting point for the construction of our model, we will first summarize the main ideas which lead to the introduction of the PB model.

The authors focus mainly on the energetic aspects of the process. The open complex formation requires the breaking of about 20 *bps*: this implies that a certain amount of energy should be concentrated in that region. Such a denaturated bubble can be obtained by heating the molecule (thermal denaturation), but this would require a greater temperature than the physiological one. Furthermore, it is known that the enzyme responsible for transcription, the RNA-polymerase, does not bring energy in the early phase of the process. It is an interesting problem to understand how such a strong *energy localization* can be achieved in DNA molecules without any expense for RNA-polymerase.

Description of the model

Peyrard and co-workers [14, 33] build initially a planar model where each strand is represented by a one-dimensional lattice of point masses corresponding to nucleotides[5] (Figure 2.2). The model is homogeneous: no distinction between the different bases, A, T, G and C is introduced. Only transversal displacements of the two complementary nucleotides in the n-th pair with respect to the molecular axis are considered. Their positions are denoted respectively w_n and v_n (Figure 2.3). An on-site *Morse potential*

$$V(w_n - v_n) = D(e^{-\alpha(w_n - v_n)/\sqrt{2}} - 1)^2 , \tag{2.3}$$

where the constants D and α are related to the depth and the width of the potential well respectively, describes the strength of the hydrogen bond linking the two bases. The choice of the Morse potential is not due to fundamental reasons. What is important is to mimic the main features of the hydrogen bonds: the interesting features of the Morse potential are the presence of a finite equilibrium

[5] Due to this strong simplification of the molecular structure, we will call the lattice units sometimes "nucleotides", sometimes "bases" : the point mass corresponds to the mass of an entire nucleotide, nevertheless we will consider bases as central interaction units.

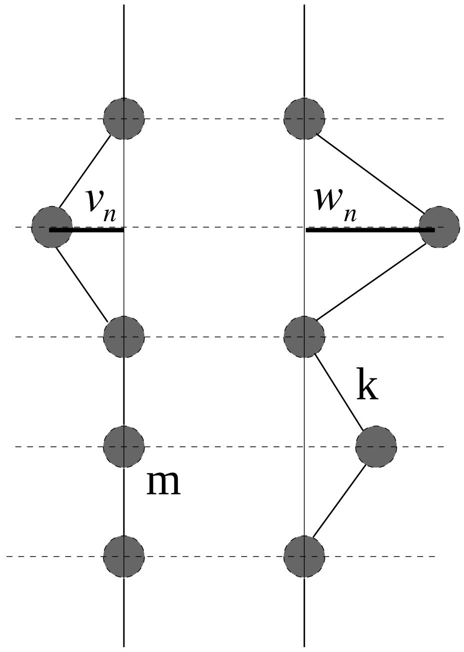


Figure 2.2: Schematic representation of the DNA molecule in the Peyrard Bishop model. In the PB model, all nucleotides are assumed to be point masses of mass m. Each of them is linked to the neighbors on the same strands by springs of constant K. The molecule has a planar geometry with transversal stretching (v_n, w_n) as unique possible motion of bases.

distance and a plateau for sufficiently large distances. These properties represent quite well the hydrogen bond, allowing, on one hand, an harmonic approximation to describe the small oscillations around equilibrium and, on the other, the description of the breaking of this bond, through the transition to the zero force regime for large base separation.

The neighboring bases on each strand are coupled through a quadratic potential with elastic constant K, which represents the stacking interaction between base pairs. This potential reads

$$W(v_n, v_{n-1}, w_n, w_{n-1}) = \frac{K}{2}\left[(w_n - w_{n-1})^2 + (v_n - v_{n-1})^2\right]. \tag{2.4}$$

The authors stress that a more convenient form for the model Lagrangian can be obtained by introducing normal coordinates

$$x_n = (w_n + v_n)/\sqrt{2} \tag{2.5}$$
$$y_n = (w_n - v_n)/\sqrt{2}. \tag{2.6}$$

By this means, the terms of the Lagrangian depending on x_n only, which are related to the in-phase motion of the two nucleotides in a pair, are decoupled.

This part corresponds just to an harmonic chain without on-site potential. It will be ignored, because it trivially gives the linear phonon excitations of the chain of centers of mass of the base pairs.

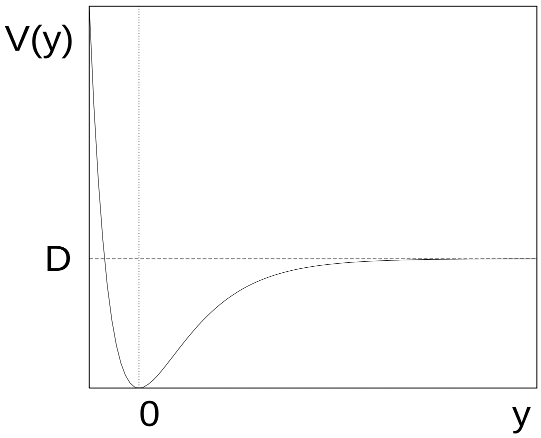


Figure 2.3: Morse potential shape. The morse potential well is characterized by two parameters, its width α and its depth D.

The hydrogen bond stretching y_n are the only variables which are considered. Therefore, each base pair is now described by a single degree of freedom. The resulting (nonlinear) Lagrangian of the PB model is

$$\mathcal{L}_{PB} = \sum_n \frac{m}{2}\dot{y_n}^2 - \sum_n D\big(e^{-\alpha y_n} - 1\big)^2 - \sum_n K\big(y_n - y_{n-1}\big)^2 . \tag{2.7}$$

Energy localization phenomena are often observed in non linear systems. Dauxois *et al.* [34] (see also [35] and [36]) have looked for localized excitations of the PB model. They have shown that it is possible to find analytically special (approximate) soliton-like solutions of the form of breathers, by applying a standard Multiple Scale Expansion technique.

2.2.2 The Multiple Scale Expansion solution of PB model

One of the most popular approaches used to determine the small amplitude envelope soliton solutions in one-dimensional scalar non linear models is the *Multiple Scale Expansion (MSE) technique*. The main idea of the MSE, as it has been developed by Remoissenet [37], is outlined, as far as its application to the PB model is concerned, in [35]. We summarize here briefly the main steps of its application[6].

[6] We also refer the reader to Chapter 3, where the MSE technique will be generalized, for a more detailed discussion.

After having introduced dimensionless variables by rescaling spatial lengths and time, and expanded the Morse potential in Taylor series, Ting and Peyrard [35] assume that the amplitude of the distortion is small and that it is an approximate envelope soliton. This is a solution of the form of a wave packet, modulated by a slowly varying envelope function, which is spatially localized. The behavior of the solution can thus be described on different time and space scales: on the smaller scales the high frequency internal vibration plays the main role, and on the longer time scales the slowly varying envelope enter into play.

If one assumes in fact that the solution has the form

$$y_n = \epsilon u_{1,n} + \epsilon^2 u_{2,n} + O(\epsilon^3) \tag{2.8}$$

where $u_{1,n}$ and $u_{2,n}$ represent the contributions at different perturbation orders, the order ϵ equation gives for the ϵ order term in (2.8) the following modulated wave packet form

$$u_1 = F_n(t)e^{qn-\omega(q)t} + c.c.\,, \tag{2.9}$$

where q is the wave number, $\omega(q)$ is the corresponding frequency of the linearized system, and F is an envelope function. The method then consist in separate the behavior of the solution on the different scales: one define the variables t_0, t_1, t_2 and x_0, x_1, x_2, which can be identified with $t, \epsilon t, \epsilon^2 t$ and $n, \epsilon n, \epsilon^2 n$ respectively, and considered as independent. This leads to rewrite the equations of motion at different orders separately, by introducing the relations

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial t_0} + \epsilon\frac{\partial}{\partial t_1} + \epsilon^2\frac{\partial}{\partial t_2} + O(\epsilon^3) \tag{2.10}$$

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial x_0} + \epsilon\frac{\partial}{\partial x_1} + \epsilon^2\frac{\partial}{\partial x_2} + O(\epsilon^3)\,. \tag{2.11}$$

Then F, which is a slowly varying envelope, is assumed to be a continuous function of the variables x_1, x_2, t_1, t_2. Note that only the slow time and space dependence of the solution is treated in the continuum approximation, while the discrete description is maintained for the fast oscillation of frequency $\omega(q)$. This is a special MSE approach, called *semidiscrete*, which allows to take into account the discreteness at least on the small scales[7].

Substituting (2.8), (2.10) and (2.11) into the equations of motion derived from (2.7) and equating terms of the same power in ϵ, one gets a set of equations in increasing powers of ϵ that can be solved sequentially.

The equation of higher order allows to find u_2 as functions of F. Finally, after a change of reference frame defined by

$$\begin{aligned} s_i &= x_i - Vt_i \\ \tau_i &= t_i, \end{aligned} \tag{2.12}$$

[7] A simpler approach is obtained starting directly from a continuum approximation of the equations of motion.

where V is the linear dispersion velocity of the system $\partial\omega/\partial q$, one finds that F is the solution of the following NLS equation:

$$i\frac{\partial F}{\partial \tau_2} + P\frac{\partial^2 F}{\partial s_1^2} + Q|F(s_1,\tau_2)|^2 F(s_1,\tau_2) = 0\,. \tag{2.13}$$

Envelope soliton solution of the NLS and final solution

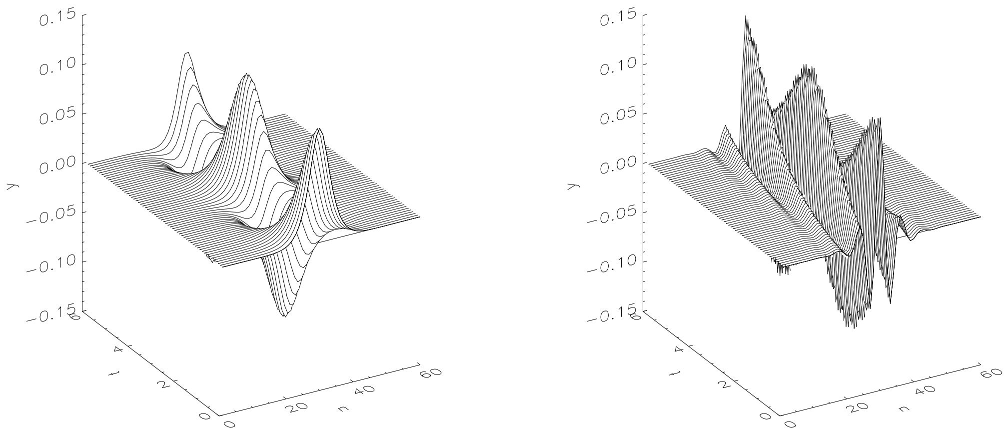
The coefficients P and Q are functions of the wave number q. They represent the dispersive and the non linear features of the model respectively. Dispersion, *i.e.* non linear dependence on q of the plane wave frequencies, tends to spread over a localized distortion F, because its Fourier components travel with different velocities. The effect of non linearity is the opposite, and their balance tends to maintain the distortion. In comparison with the linear Schrödinger equation, the last term in (2.13) represents the effect of a potential well $Q|F|^2$ that depends on the function F itself: this potential becomes deeper and therefore more attractive when the function F becomes narrower. This would tend to the collapse of F, if the dispersion term were absent. The competition between dispersion and non linearity in the NLS equation is such that the two effects of dispersion and collapse balance each other allowing the existence of localized solutions of well defined shape.

If the product PQ is positive the NLS equation admits an envelope soliton solution of the form (*cf.* Appendix A):

$$F(s_1,\tau_2) = A\,\mathrm{sech}\Big(\frac{1}{L_e}(s_1 - u_e\tau_2)\Big)\;\exp\big(i\frac{u_e}{2P}(s_1 - u_c\tau_2)\big)\,, \tag{2.14}$$

where u_e is the envelope velocity (in the reference frame defined by (s_1,τ_2)), u_c the carrier velocity, $A = \sqrt{(u_e^2 - 2u_eu_c)/2PQ}$ and $L_e = 2P/\sqrt{(u_e^2 - 2u_eu_c)}$ respectively the envelope amplitude and width.

Equation (2.8) gives the complete approximate solution, whose main contribution $\epsilon u_{1,n}$, written in terms of continuous variables, is, as given in Equation (2.9), a plane wave modulated by an envelope F with the sech shape, moving with velocity $V + \epsilon u_e$, with internal oscillation of frequency $\omega + \epsilon u_e(V + \epsilon u_c)/2P$ and wave vector $q + \epsilon u_e/2P$. The solution is, as anticipated, a breather: a sketch of its behavior for two different values of q is given in Figure 2.4. The envelope oscillation is asymmetrical, due to the asymmetry of the Morse potential well, corresponding to a positive mean value of the stretching y_n. These localized oscillations are well stable in the small amplitude limit, in the sense that, provided that their amplitudes are sufficiently small for the approximation to be valid, they can propagate along the chain for long times with a very small loss in energy and without changing their shape.



(a) (b)

Figure 2.4: Time evolution of the solution (2.8) for the PB model, having used $\epsilon = 0.1$, $u_e = -0.1$, $u_c = 0.1$. **(a)**: for $q = 0.1$ ($P \approx 0.023$, $Q \approx 1.87$, $V \approx 0.056$); **(b)**: for $q = 1.2$ ($P \approx 0.025$, $Q \approx 2.00$, $V \approx 0.050$). The internal oscillation is characterized by wave number $q + \epsilon u_e/2P$ and frequency $\omega + \epsilon u_e(V + \epsilon u_c)/2P$: note that it is faster in the case of greater q. The two solutions move along the chain with a small velocity: their translation cannot be visible within the small time range displayed.

2.2.3 Thermal breathers as precursors for transcription

The PB moving breathers are of very small amplitude, so that they can probably not be considered as traveling activation distortions carrying activation signals from enhancers to promoter sites of the type suggested in Chapter 1. Their amplitudes are also small with respect to that of the open complex. Peyrard and co-workers suggest anyway that they could be considered as precursors for the transcription bubble formation. Several results contribute in fact to indicate a possible role of breathers in this biological function.

Thermal denaturation simulations in the canonical ensemble

Dauxois *et al.* introduce in [33] the analytical and numerical investigation of the dynamics and thermodynamics of the denaturation process in the context of the PB model. Denaturation, obtained by heating the DNA, is essentially characterized by the formation of local bubbles, whose extensions grow with temperature, up to the complete separation of the two strands. Bubble formation and bubble growth are the questions investigated in [33]. The dynamics of the model in contact with a thermal bath is analyzed, for temperatures approaching the melting point, through molecular-dynamics simulations. The fixed temperature canonical ensemble is simulated by the addition of a single variable which regulates the energy flow, according to the Nosé Hoover method [38, 39, 40].

In agreement with experimental results on DNA "breathing" [41, 23], the authors find for low temperatures the formation of small oscillating openings, well localized on a few lattice sites. This shows that the system is far from a situation of energy equipartition - in the small to medium time range. These oscillations are likely to be the small amplitude breather oscillations obtained by the MSE expansion, even if this cannot be proven rigorously on the basis of the canonical simulation results. For increasing temperatures, the small oscillations grow in amplitude and width, and combine to lead to a complete strand separation, as in real DNA.

Energetics of breather interaction

In a thermalized chain, breathers may form spontaneously by energy localization. Furthermore, they have the tendency to grow. In order to confirm this suggestion, obtained by the canonical simulation of denaturation, it has been shown [33] that two breathers interact in collisions in such a way that the energy exchange between them tends in average to favor the growth of the larger excitation. Moving breather excitations can even be created in a PB chain by modulational instability, *i.e.* through the instability effect induced by a small perturbation on a plane wave solution of the model. These breathers show again a tendency to grow by collisions, increasing their amplitude and decreasing their width [42], until they are trapped and stopped by discreteness effect [43].

A possible role for breathers in the open complex formation

All these results suggest that breather excitations can play the role of energy carriers in biological functioning, when an energy focusing on special sites is requested. Along these lines, a last interesting indication is given by another property of these solutions: not only the discreteness of the lattice acts as a trap, but they can also be trapped by local inhomogeneities of the chain, namely by a local change in the coupling constant K along the strands. Such a change could be induced in real DNA by a bending of the helical axis, because bases that are inside the bend are brought closer to each other and bases that remain outside are moved apart. Even if the effect on the two chains is different, in a first approximation the overall effect on the relative distance y_n can be represented as a local modification of the stacking coupling. Approximate breathers of this type, traveling toward the inhomogeneity, can be trapped if the inhomogeneity itself is strong enough [35].

This allows the authors to put forward a hypothesis for a possible mechanism of the open complex formation. Since we know that the RNA-polymerase bends the promoter region, as described in Chapter 1, one can suppose that this bend causes a local change of the coupling constant K. This means that it is possible to mimic the action of the bound RNA-polymerase with the introduction of a local inhomogeneity that acts as a breather trap and collects in that region enough energy to allow bubble formation.

Summarizing, the PB model provides breather-like approximate solutions, characterized by an oscillatory motion (opening/closing of the bonds), localized in small regions of the chain and propagating along the chain itself. These solutions are of small amplitude. Therefore, they cannot correspond to the bubble formation in which we are interested. Nevertheless they have some interesting properties from a biological point of view:

1. *They can be produced by spontaneous thermal energy localization,*
2. *By nature, they tend to collect energy,*
3. *They can be trapped by local inhomogeneities of the chain.*

This mechanism can be an explanation for the bubble formation in the context of transcription initiation processes.

2.2.4 Effects of the specific DNA sequence

Before proceeding, it is interesting to mention the possibility of including inhomogeneities in DNA models of various types, even if this argument will not be treated in this work. Until now, we have not considered the effect of the base sequence on DNA dynamics. In reality, the mechanical properties of each DNA region depend strongly on its specific base content: not only the relative strength of AT and CG base pairs coupling, but also the stacking force, the helix form, the molecule bendability, the degree of supercoiling and so on. The effect of the base sequence can be crucial for what concerns the dynamical behavior of localized movable soliton-like distortions, as it has been pointed out by Salerno [44]. He worked on an inhomogeneous version of the Yakushevich model [11], a model that takes into account the rotation of each base around its strand. This model leads to a description of the DNA through a sine-Gordon equation, so that one gets the exact soliton solutions of this equation as traveling distortions. Introducing in the model a real promoter sequence by distinguishing double and triple hydrogen bonds, Salerno has shown that the resulting effective potential "landscape" can act on the motion of a soliton solution by accelerating, decelerating or reflecting it, depending on the soliton initial velocity. This effect could have an important role in the beginning of the elongation phase, where the transcription bubble starts to move in the reading direction of the coding sequence.

For what concerns the PB model, some attempts have been made in this direction mainly in the context of denaturation studies [45, 46], in order to point out the sequence dependence of the denaturation temperature and of the denaturation curve shape. In particular Cule and Hwa [46] obtain good results in reproducing the characteristic "multistep" fine structure of melting curves observed in real experiments, by considering random inhomogeneous sequences. They also give a theoretical explanation for this structure[8]. We will not deal with this topic in the present work, but the introduction of real sequences by varying the intra-base coupling constant represents a good way to improve DNA models which has to be kept in mind. Moreover, this represents a link to another interesting research field concerning the statistical properties of DNA sequences.

2.3 Motivations for the proposal of a new geometrical model

PB model limitations

The PB model is a good starting point because it is able to reproduce some interesting features of the energetics and dynamics of DNA. However, this model is in fact too simplified. As we have stressed in Chapter 1, DNA dynamical behavior cannot be properly described without considering the effects arising from the helical structure. Because of the coupling between base pair opening and twist (See Section 1.2) these effects become crucial in transcription and transcription initiation, when a strong coiling of the two strands could lead to a strong restriction to the bubble formation and to the bubble motion. Therefore, geometrical constraints have to be considered in order to describe the dynamical features involved in transcription in a more realistic way. It is impossible to take fully into account this geometrical information (and thus its physical consequences) in the context of the the PB model, because of its intrinsic planar nature.

Our aim is the investigation of the specific role of the helicity in transcription initiation, by mean of a simple model.

Interests in the introduction of twist

The direct modeling of the DNA helicity.

We can imagine to obtain the "transformation" of an ideal planar DNA ladder into the double helix structure simply by the introduction of a rotation of each ladder step, representing a base pair, of a fixed angle with respect to the previous one. This rotation angle is exactly what is called the *twist* angle in DNA,

[8] More recently, Campa and Giansanti [47] have used again an inhomogeneous version of the PB model which takes into account the real sequence to reproduce the denaturation curves for small real DNA chains of about $20bps$ with good approximation.

and represents the natural parameter to introduce in order to take into account molecular helicity.

Through the introduction of the twist parameter we can consider the true molecular structure and analyze all the dynamical features directly related to the helicity, as the untwist induced by the bubble opening, but also local transitions from B-DNA to the left-handed Z-DNA, effects in circular molecules and free ends molecules, or supercoiling effects in straight DNA.

Twist angle is a pertinent quantity for biological understanding.

Biological experiments have shown the relevance of this geometrical parameter for DNA dynamics. In fact, twist is directly implied in the important correlations between structure and functioning in DNA we are interested in. We know from Section 1.2 that twist angle is strongly coupled to opening, and that it can strongly influence the formation and the motion of the transcription bubble. The opening of the two strands is impossible without a local untwist. To open regions where the twist is particularly high may result impossible; the over-twist produced in the vicinities of an open bubble must be in fact eliminated by special enzymes during transcription in order to prevent the bubble from collapsing.

Twist can be introduced as a local degree of freedom maintaining the one dimensional lattice model, without the need of three dimensional description. By coupling local twists and hydrogen bonds stretching in an appropriate, geometrical way, we will be able to obtain a model that contains the fundamental features of the helical structure and that mimics quite well its possible dynamical deformations.

Twist angle includes indirectly other DNA structural deformations.

We stress that the twist angle can represent indirectly other DNA structural deformations. In fact, there exists strong geometrical links between the various helix deformations in the three dimensional configuration - bending, supercoiling - as well as in the local helix shape - stretch of the hydrogen bonds, changes in twist etc. These quantities are all related to each other [26, 1]. It is remarkable that the bending of the chain implies that, in the same region, the twist angle tends to *decrease* [19]: this twist decreasing has to occur when the RNA-polymerase binds to the molecule. Furthermore we know that a decrease of twist is a necessary condition for the opening of the chain: opening may be directly induced by the bending of the promoter region due to RNA-polymerase interaction.

The main reason why we introduce the twist angle is that it describes the molecule helicity. We consider it as the degree of freedom which is coupled with the opening deformations in the most direct way. At the same time, we will keep in mind that twist could depend on three dimensional deformations like bending and supercoiling. We could perhaps, in the future, include indirectly further informations on some three dimensional properties by considering special constraints on twist.

2.4 Presentation of the *twist-opening* model

We propose [16, 15] a model based on the introduction of the helical structure into the description of the molecule, by the addition of one more degree of freedom to the hydrogen bond stretching, namely the *twist angle*. Biological experiments have shown the relevance of this geometrical parameter for DNA dynamics. We will investigate if the results and the hypotheses made on the basis of the planar PB model remain valid in the context of a more accurate description of real molecular structure. In particular we will look for an extension of PB breathers in the two degrees of freedom model: it is in fact interesting to understand if an oscillating opening solution can exist even in a more complex model, and which kind of radial distortions should accompany it.

Moreover, the way we will describe DNA corresponds to a quite different approach in modeling. We will focus on a quite rigorous reproduction of the helix geometry. We will pay more attention to the structural and geometrical features of the model instead of its mathematical properties when used for the description of a chosen phenomenology. This allows the construction of a model whose interest could be not only in the transcription initiation studies, but more generally in describing other DNA dynamical features as *e.g.* local transitions from B-DNA to the left-handed Z-DNA, specific behavior, during transition, of circular molecules with respect to molecules with free ends, supercoiling effects in straight DNA, and secondary effects of bending.

2.4.1 Construction of the model

The point masses and the radial degree of freedom

In designing such a model, each group made of a sugar ring and its connected base (without distinction between the different base types) will be considered as $\boxed{m}$ a single, non-deformable object, with mass $m \approx 300$ *a.u.m.* which is an averaged estimation of the nucleotide mass.

Being interested in the transcription and transcription initiation processes, our first task is to describe DNA openings, as in the PB model. However, the model does not attempt to describe the acoustic motions of the molecule: then we consider only the stretching of the base pair distance. This amounts to fixing the center of mass of the base pair, *i.e.* the two bases in a pair are constrained to move symmetrically with respect to the axis of the molecule. So, to describe the stretching of a base pair we only need a single degree of freedom per base pair, say the distance from the molecular axis of one of the two bases. We will thus introduce one radial variable related to the motion of the bases along the diameter $\boxed{r_n}$ that joins their attachment points to the helicoidal backbone, r_n (Figure 2.5).

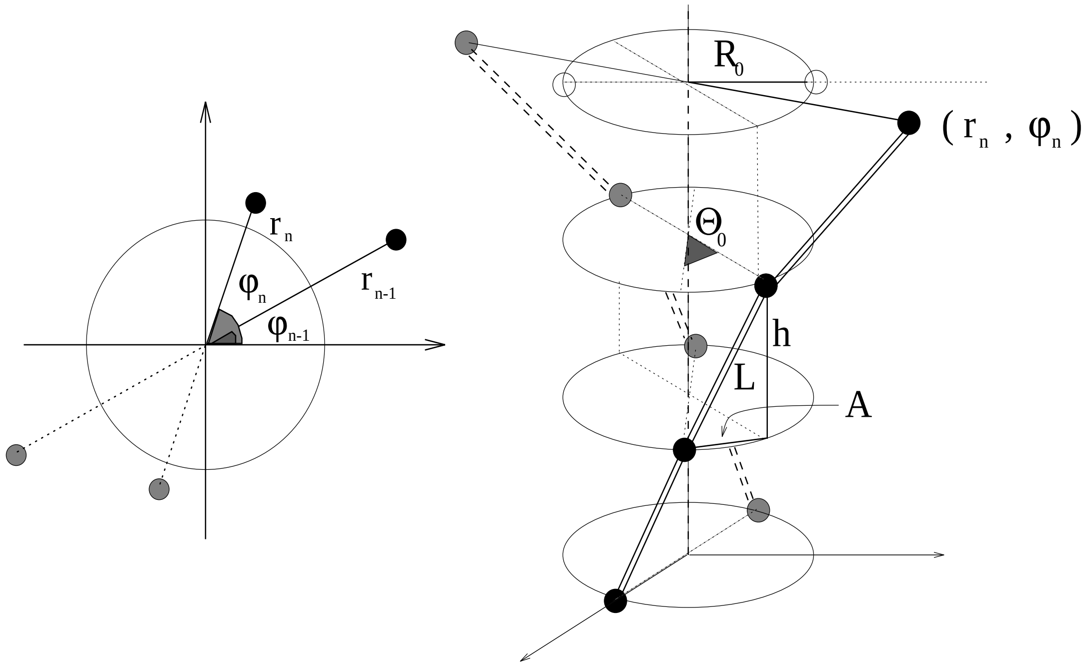


Figure 2.5: The choice of variables (left) and the schematic view (right) of the model. The n-th base pair is represented by its radial and angular displacements (r_n, φ_n) with respect to a fixed external reference frame. Neighboring base pairs are linked along the two strands by elastic rods with equilibrium length L. The constants R_0, Θ_0 and h are the geometrical parameters of the model, with values deduced from the real B-DNA geometry.

The hydrogen bond potential V_m

In each pair, the two masses representing the bases have to be coupled in order to reproduce the effect of hydrogen bonds in real DNA. As in the PB model, we describe the hydrogen bonds by a *Morse potential* V_m, with depth D and width α. $\boxed{D}$ $\boxed{\alpha}$

Furthermore, in order to take into account the true geometrical configuration of the DNA molecule, we introduce explicitly the non zero equilibrium molecular diameter R_0 which corresponds to approximatively 10 Å for the B-DNA. The $\boxed{R_0}$ Morse potential will then depend on the stretch variable $r_n - R_0$, which represents the displacement of the base pair with respect to its equilibrium configuration:

$$V_m = \sum_n D\left(e^{-\alpha(r_n - R_0)} - 1\right)^2 . \qquad (2.15)$$

The twist degree of freedom

A direct way to introduce the additional twist motion is to allow the two bases in each pair to move in the base pair plane instead of constraining them on a line.

It is useful to choose a polar coordinate system.

$\boxed{\varphi_n}$ We introduce then a second degree of freedom per base pair: the angle φ_n defined as the angle made by the base pair axis (the ladder "step") and a reference direction given by an external reference frame. The twist is then defined by the $\boxed{\theta_n}$ difference between neighboring base pair angles, $\theta_n = \varphi_n - \varphi_{n-1}$ (Figure 2.5).

The stacking interaction W_{rods}

In order to investigate the dynamical effects arising from the DNA helicity, we have first of all to reproduce its equilibrium structure: therefore, the coupling between the two degrees of freedom has to be chosen in order to stabilize it in its zero energy state. To obtain the right coupling between radii and angles, we can, in fact, explicitly refer to the helicoidal geometry.

The helical shape of the molecule is essentially due to the competition among three forces: the *hydrophobic interaction*, the *electrostatic repulsion* and the *backbone rigidity*. The first interaction arises from the hydrophobic character of the bases: this tends to eliminate water from the core of the molecule by bringing the neighboring base pair planes closer. The electrostatic repulsion between neighboring base planes has the opposite effect. We combine these two forces in an unique *stacking effect* by fixing the base-plane distance in the model. Moreover, bases along the same strand are connected by a sugar phosphate backbone segment which is quite rigid, so that it separates the external ends of the base pairs essentially by a fixed length: to bring base pairs closer it is necessary to rotate them, inclining the backbone in the helical structure [19].

To reproduce the competition between these two opposite effects we allow the $\boxed{h}$ base pairs to move on planes whose distance h is fixed to the approximate B-DNA base pair step, $h \approx 3.4$ Å. We describe the link between neighboring bases along the strands, due to the rigid backbone segments in between, by elastic rods with $\boxed{K}$ stiffness K. We fix the elastic rod length at equilibrium to a value L greater than $\boxed{L}$ h, in such a way that the rods, in the equilibrium position, have to be inclined (Figure 2.5).

The resulting geometry has to be such that the corresponding rotation of $\boxed{\Theta_0}$ neighboring base pairs is $\Theta_0 \approx 36°$, *i.e.* the B-DNA equilibrium twist value. This is obtained by fixing the constant L to

$$L = \sqrt{h^2 + 4R_0^2 \sin^2\left(\frac{\Theta_0}{2}\right)} > h \tag{2.16}$$

which, in fact, is its approximate value for B-DNA.

In a generic distorted configuration, the elastic rods have length equal to

$$L_n = \sqrt{h^2 + r_{n-1}^2 + r_n^2 - 2r_{n-1}r_n\cos(\varphi_n - \varphi_{n-1})}\,. \tag{2.17}$$

Correspondingly, we will introduce in the Lagrangian of the system quadratic terms W_{rods} depending on the difference between this length and L:

$$W_{rods} = \sum_n K\Big(L_n - L\Big)^2 . \qquad (2.18)$$

Strand curvature term W_{cur}

However, because of the invariance of L with respect to the sign of Θ_0, the elastic rods rigidity is not sufficient to guarantee the correct helicoidal shape. There is in fact no constraint for the rotation to be clockwise or anticlockwise, so that we could have a zig zag behavior with rods randomly inclined left or right-handed. To obtain a true helical shape we have to add a three-body curvature term

$$W_{cur} = \sum_n G_0 \left(\varphi_{n+1} + \varphi_{n-1} - 2\varphi_n\right)^2 . \qquad (2.19)$$

which imposes a continuity between the differences $\varphi_n - \varphi_{n-1}$ and $\varphi_{n+1} - \varphi_n$. It essentially constrains each twist angle to have at equilibrium the same sign than the previous one. We denote G_0 the elastic constant related to this interaction term. $\boxed{G_0}$

The resulting Lagrangian of the model

With the final Lagrangian [16] which reads

$$\begin{aligned}
\mathcal{L} &= E_{kin} - V_m - W_{rods} - W_{cur} \\
&= \sum_n \big(m\dot{r_n}^2 + m{r_n}^2\dot{\varphi_n}^2\big) - D\big(e^{-\alpha(r_n - R_0)} - 1\big)^2 \\
&- \sum_n K\Big(\sqrt{h^2 + r_{n-1}^2 + r_n^2 - 2r_{n-1}r_n\cos(\varphi_n - \varphi_{n-1})} - L\Big)^2 \\
&- \sum_n G_0 \left(\varphi_{n+1} + \varphi_{n-1} - 2\varphi_n\right)^2 . \qquad (2.20)
\end{aligned}$$

our model is completely determined. Nevertheless, a correct choice for the model constants has to be done. This choice can be an hard task, especially for K and G_0, as we will see, but it is of fundamental importance in order to represent as well as possible the DNA dynamical features. For this problem, we refer the reader to the Section 2.5 where it will be considered on the basis of experimental indications.

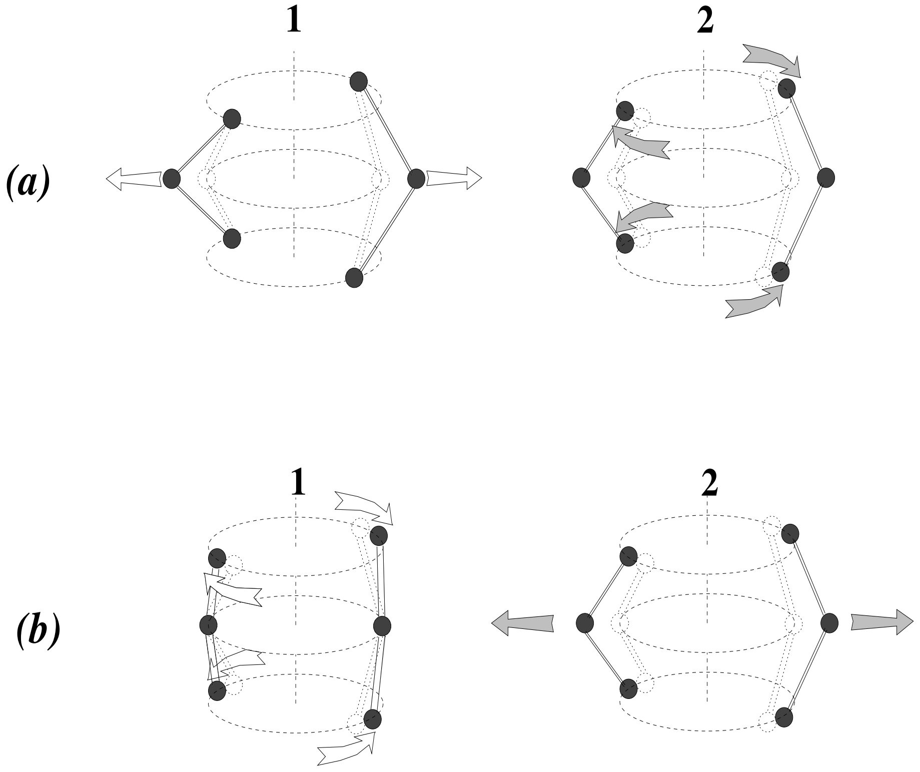


Figure 2.6: Schematic illustration of the coupling between angles and radii that results from the geometrical constraints. **(a)**: the stretching of a base pair tends to untwist the two neighboring pairs; **(b)**: the rotation of two neighboring base pairs toward a common plane pull a central couple of bases out of the cylinder of radius R_0.

2.4.2 Some straightforward consequences on twist-opening coupling

At this point, we have almost all the elements to try our model for DNA dynamics simulations. But before, it is worth to verify if its behavior is the same as the desired one.

For the proposed model, we can already say that the equilibrium state is a *helix*. As a consequence of the imposed geometrical constraints, we obtain a *good coupling* between the two degrees of freedom: the stretch of the hydrogen bonds and the untwist of the molecule are strongly correlated. On one hand, trying to stretch a single base pair, the neighboring pairs are pulled by the rods in such a way that the twist angles tends to decrease (as expected from considerations in Chapter 1). On the other hand, if instead we rotate the neighboring base pairs in opposite directions bringing them towards a common vertical plane, the rigidity of the backbone rods acts on the direction of pushing the central base out of its equilibrium position, breaking the hydrogen bonds. Therefore, the stretch of the hydrogen bonds and the untwist of the molecule are correlated as we would like. Figure 2.6 illustrates this coupling mechanism.

Numerical check of the coupling

The correctness of the coupling between stretch and untwist can be easily verified by numerical integration of the equations of motion with particular initial conditions. Figure 2.7 and Figure 2.8 report the time evolution of variations of the radial (r_n), angular (φ_n) and twist (θ_n) variables compared to their equilibrium values, for a chain of 200 bps. We show their time evolution, obtained by numerical integration of the equations of motion derived from the Lagrangian (2.20). At the final time the radii oscillate around a new configuration: to clarify this dynamical behavior, we superimposed in the right radial graphs the configurations corresponding to several close times, namely $t = (7800 + 20\,i)$ time units with i running from 1 to 30 ($1\,t.u. = 1.02\;10^{-14} sec$).

The chain is initially at rest except in a small region. In Figure 2.7, the initial condition corresponds to the opening of a single base pair: its displacement is of 1 Å and corresponds approximatively to the opening length of one hydrogen bond [22]. Angles (and twists) are initially at rest. The integration of the equation of motion induces then a decrease of the twist in correspondence to the stretched region. A rotation of all the upstream angles φ_n of about $+0.012$ and of all the downstream ones of about -0.012 in fact arises, and tends to stabilize by elimination of the extra-twist from the two free chain boundaries. Note that the radii start to oscillate in the distortion region. The oscillation is asymmetric: this is due to the asymmetry of the nonlinear Morse potential, that favors base pair opening with respect to radial compression.

In the second case, Figure 2.8, we start with an angular distortion that corresponds to a twist decrease of about 0.012 on site $n = 100$. This corresponds approximately to the final twist configuration of the previous case. The system initially (on time of the order of $20\,t.u.$) show its tendency to open in correspondence with the untwisted region. The radial stretch extends up to 0.005 Å. Then, the radii start to oscillate and the oscillation stabilizes between amplitudes ± 0.002 Å. In this case, the amplitude is small and corresponds to an oscillation on the bottom of the potential well: for this reason, the asymmetry is small. The radial dynamics and the rod coupling have in this case a back effect on the helical angle: the initial twist is eliminated by translation towards the two chain ends and the equilibrium configuration is restored, compatibly with the smallness of the induced radial oscillation and with the fact that it corresponds to a zero mean displacement.

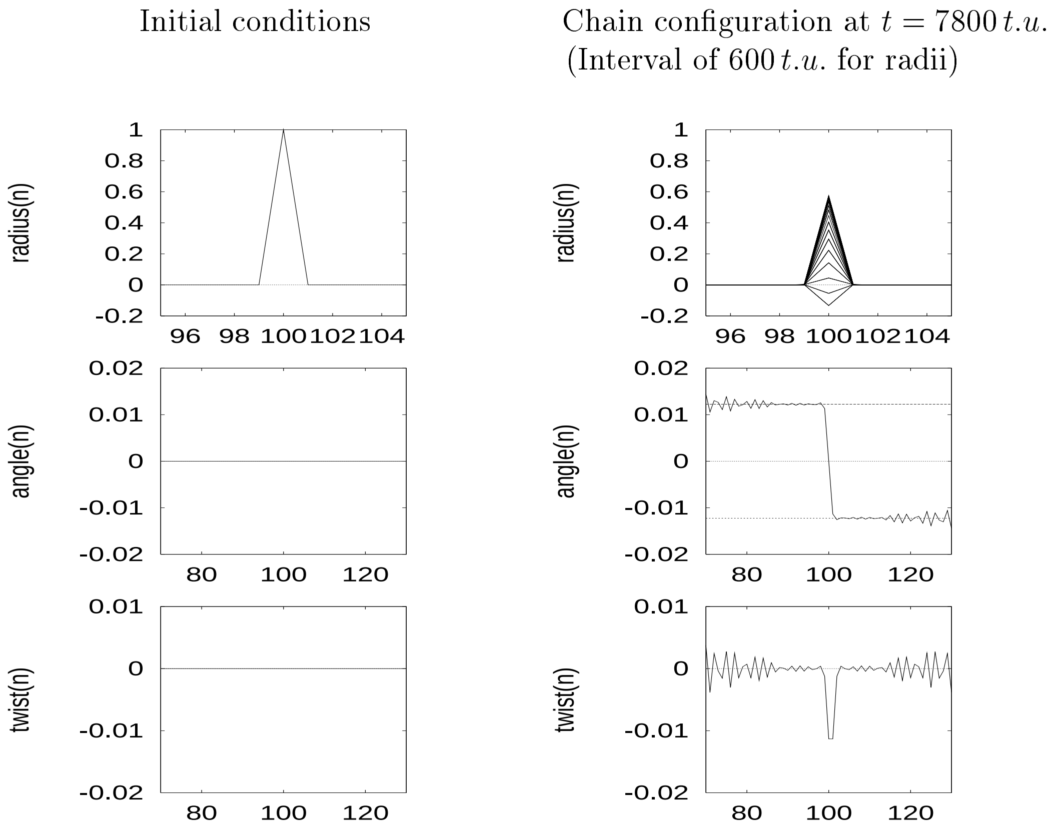


Figure 2.7: Simulations of the *twist-opening* model with simple initial distortions. We integrate the complete equations of motion of the system with, as initial condition ($t = 0$), a local radial stretched region. The time evolution of the system is shown: the first column displays the initial condition, the second column represents the system configurations after integration for 7790 time units. In the radial graph we superimpose 30 subsequent configurations with a time interval of about $20\,t.u.$

Simulations presented in this section are performed with the following set of parameters: $D = 0.04\ eV$, $\alpha = 4.45$ Å^{-1} and $K \approx 0.04\ eV$Å^{-2}. These choices will be justified in the following section. The choice of the G_0 value is, as we will see, more difficult. We used here $G_0 = KR_0^2$. Anyway, the results we presented are very slightly affected by changes of G_0 in the range $0.1 \div 10\ KR_0^2$.

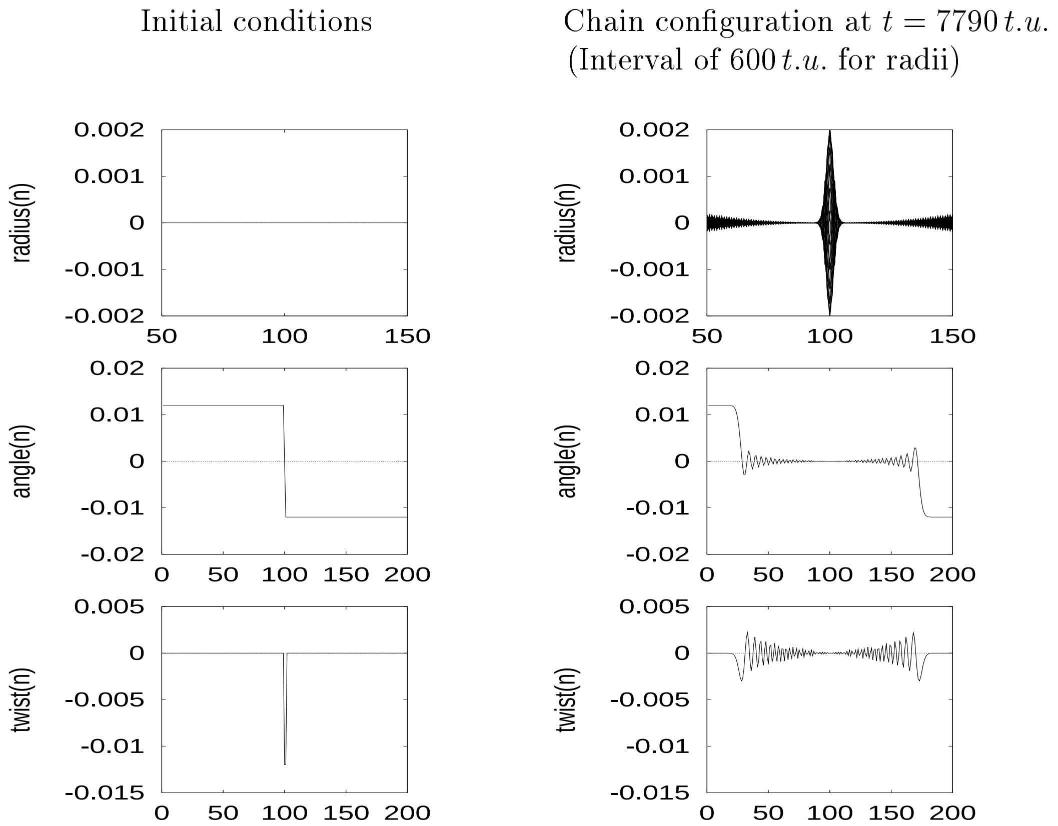


Figure 2.8: Simulations of the *twist-opening* model with simple initial distortion. We integrate using as initial condition, an untwisted region. The first column displays the initial condition, the second column represents the system configurations after integration for 7790 time units. In the radial graph we superimpose again 30 configurations with $20\,t.u.$

2.5 Determination of the model constants

Several constants appear in the Lagrangian of the model. Being interested in a simplified description of the chain dynamics, we do not need a very precise determination of these constants. Then, we will look for a realistic estimation of their order of magnitude, to fulfill the aim of describing the main characteristic of the real molecular motions.

Some of the involved constants are simply geometrical parameters. Their orders of magnitude are known with a sufficient precision: that is the case of $R_0 \approx 10$ Å, $\Theta_0 \approx 36°$ and $h \approx 3.4$ Å [19]. The same could be said for the (averaged) base mass $m \approx 300$ $u.m.a.$ [18].

2.5.1 Morse potential parameters

The determination of the parameters describing the interaction forces is a delicate problem due to the lack of precise experimental results. In the Morse potential term, the D constant corresponds to the depth of the Morse well, and characterizes the strength of the hydrogen bond. It is thus related to the energy needed to break the bond and to separate the two bases. The parameter α^{-1} represents the width of the potential well, *i.e.* it gives an indication on the distance between bases at which the pair denaturate.

The DNA dissociation energy value is known to correspond to some $k_B T$, where k_B is the Boltzmann constant and T the (room) temperature. Recently, partial results of an experimental manipulation on single DNA chain [48] gave an estimation for this DNA base pairs denaturation energy. It has been done on the basis of an hysteresis cycle in the force-elongation diagram for a molecule twisted and stretched by pulling the two ends. According to these results, a good estimation for the dissociation energy D is $1.36\, k_B T_{den}$. This gives, for a denaturation temperature of approximately $T_{den} = 350K$, a dissociation constant $D = 0.04\ eV$. Furthermore, this value is in good agreement with the value adopted for the PB model [33] and we will make the same choice.

For what concerns the other parameter involved in the Morse term, α, it is not simple to fix a very precise value: the force dependence on the distance between the bases in a pair is not only related to the hydrogen bond interaction, but also to a complex series of electrostatic interactions between the various atomic groups in the base pairs and in the backbone. In a first approximation, it is possible to obtain its order of magnitude considering the fact that the average hydrogen bond length is of about 2.9 Å, and that the relative lateral motion of the bases, which should correspond to the oscillation in the Morse potential well, is estimated of the order of magnitude of 1 Å [22]. Furthermore, the length of H-bonds are seldom observed larger than 3.3 Å[49]. The hydrogen bond length which corresponds to the denaturation point should thus be fixed approximatively just under 4 Å. The Peyrard Bishop choice is made[9] by considering the value at which the denaturation corresponds to the 95% of the base pairs: this happens for a bond stretch of about $5 \ln 2/\alpha$. To fix α at the value 4.45 Å^{-1} corresponds to a bond stretch of about 0.8 Å up to a total length of 3.7 Å, with an extension of about 1.3 times its equilibrium value.

We will adopt also in our model the PB Morse parameters $D = 0.04\ eV$ and $\alpha = 4.45$ Å^{-1} [33]. It is worthwhile to mention that, for the PB model, the correctness of these choices is also confirmed *a posteriori* by the numerical results for the model denaturation curves obtained by molecular dynamics simulations [33, 50].

[9] Note that $y = \ln 2/\alpha$ corresponds to the inflection point of the Morse potential.

2.5.2 The elastic constants

Two last parameters remain to be chosen: first the curvature constant G_0, related to the rigidity of the backbone, which describes the strength of the force that keeps the neighboring rods aligned on a smooth helix, and second the elastic constant K which characterizes the elasticity of the inclined rods between two neighboring bases along a strand. Their choice is much more difficult, because of the lack of adequate experimental results.

The G_0 constant

A choice of G_0 based on direct experimental measurements seems impossible at present. The main difficulty arises from the fact that the corresponding Lagrangian curvature term is zero for uniform twist distortions. Experiments dealing with uniform twist do not give therefore appropriate data to be compared with this term: the twist is never strictly uniform in experiments, because it varies slightly as a function of the bases so that its local value depends on the specific sequence; anyway, the direct measurements on single chains performed in todays experimental works [5] give averaged measures such as the total molecular length or the total twist excess. Resulting data cannot be compared with local distortion terms.

Experimental results concerning the specific rigidity of a single backbone segment are on the other hand not available. To our knowledge, results from molecular dynamics simulations on all-atoms DNA models are also not provided, even if they are possible in principle: an estimate of the constant G_0 could be done in fact from the energy of the interface between a B-DNA and a Z-DNA, the right handed helix. It should be possible by imposing a given distortion that excites the energy terms in G_0 or K and then relax the structure and computes its energy. Doing this for a series of deformation would give the energy as a function of the deformation and then a fit with the model Hamiltonian could be used to determine the parameters. However, this would require very heavy computations.

Having fixed the other constants, we can check if the G_0 parameter could be fixed by performing a numerical simulation of the model denaturation. It is quite evident that the denaturation has to be affected by the force by which the motion of each backbone rod is linked to the neighboring rods, *i.e.* the rigidity of the two backbone strands with respect to bending.

Numerical experiments on denaturation require heavy simulations, in order to obtain a constant temperature equilibrium. To perform simulations with many different parameters is therefore quite cumbersome. We will present preliminary results of such experiments in Chapter 4. For the moment, where a choice is needed, we will refer to different G_0 values. Furthermore, we will see (Sections

2.6 and 3.4) that, for what concern the kind of distortions we will discuss for our model, the relevance of G_0 is rather small.

The K constant

We then need a correct choice of the K parameter. Note that a comparison with the PB staking interaction constant is not useful, because it refers just to the staking interaction, and within the frame of a completely different geometry of the model.

The parameter K can be obtained by considering harmonic terms in the small amplitude approximation of the rod term W_{rods} of the Lagrangian (2.20). Considering in fact small displacements with respect to the equilibrium configuration, $r_n - R_0$ and $(\varphi_n - \varphi_{n-1} - n\Theta_0)$, the elastic term can be approximated as follows (see also Section 2.6):

$$
\begin{aligned}
W_{rods} \approx\; & K \sum_n \frac{R_0^4 \sin^2\Theta_0}{L^2}(\varphi_n - \varphi_{n-1} - \Theta_0)^2 \\
+ & K \sum_n \frac{R_0^2(1-\cos\Theta_0)^2}{L^2}(r_n + r_{n-1} - 2R_0)^2 \\
+ & K \sum_n \frac{2R_0^3(1-\cos\Theta_0)\sin\Theta_0}{L^2}(\varphi_n - \varphi_{n-1} - \Theta_0)(r_n + r_{n-1} - 2R_0)\,.
\end{aligned}
\tag{2.21}
$$

The functional dependence on the displacements for the elastic rod Lagrangian terms in (2.21) are not standard harmonic terms: even if the rod coupling is actually a simple elastic (harmonic) coupling between the bases, the Lagrangian depends nevertheless on the rod length that has to be calculated as a distance in the three-dimensional space, which is a nonlinear function of $(r_n - r_{n-1})$ and $(\varphi_n - \varphi_{n-1})$, due to the molecular geometry. However, the first term $\sum_n(\varphi_n - \varphi_{n-1} - \Theta_0)^2$ is of the standard type, and represents exactly the torsional elastic energy term for the model in the case of fixed radii ($r_n = R_0$ for all n), which is the total potential energy if we impose also that homogeneous distortions $(\varphi_{n+1} + \varphi_{n-1} - 2\varphi_n)^2$ are zero for all n, *i.e.* when all the other potential terms are constrained to be zero.

This term can be directly compared with the experimental results obtained for the torsional rigidity of the DNA molecule. The corresponding Lagrangian term to consider is the one obtained by describing the DNA chain simply as an elastic cylinder of length ℓ. The torsional rigidity of the DNA double chain is usually given in literature in terms of the twist persistence length C introduced by the following torsional energy definition:

$$
\mathcal{E}_t = k_B T \frac{C}{2}\int_0^\ell ds\Big(\frac{\partial\Theta(s)}{\partial s}\Big)^2 \tag{2.22}
$$

where $\Theta(s)$ is the local twist variation with respect to the equilibrium configuration. In the discrete scheme and rewritten in our variables $\partial\Theta(s)/\partial s$ becomes $(\varphi_n - \varphi_{n-1} - \Theta_0)$ and the previous relation reads

$$\mathcal{E}_t = k_B T \frac{C}{2} \frac{1}{h} \sum_n (\varphi_n - \varphi_{n-1} - \Theta_0)^2 \,. \tag{2.23}$$

We can identify this torsional energy with our potential term in (2.21) obtaining

$$K = \, k_B T \, \frac{L^2 C}{2hR_0^4(\sin^2\Theta_0)} \,. \tag{2.24}$$

The value $C \approx 750$ Å is well established in the biological research [51] and well confirmed in the context of recent experimental results on the DNA elasticity properties [6, 52, 53, 54]. With this value we obtain

$$K \approx 0.04 \; eV \text{Å}^{-2} \,. \tag{2.25}$$

However, further comments can be made regarding the hypotheses chosen in the previous derivation of the value of K.

Remark 1: *the constant radius approximation.*

The selection of only the harmonic terms $(\varphi_n - \varphi_{n-1} - \Theta_0)^2$ for the determination of the elastic constant K amounts to neglecting the coupling of radii and angles, or, in other words, the ability of DNA torsional deformations to increase or decrease the molecular diameter. But it is possible to evaluate numerically the elastic constant of the model by energy minimization techniques, constraining a chain with N sites to have a small excess $\Delta\theta$ in its global twist (*e.g.* with $\Delta\theta \leq 10^{-4}$), and looking for the energy dependence on this angular distortion. The comparison with the theoretical evaluation of K given before represents then a validity test.

Relation (2.23) gives the energy as a function of the squared total added twist $N \sum_n^N (\phi_n - \phi_{n-1} - \Theta_0)^2 = (\Delta\theta)^2$. It can be rewritten as

$$\mathcal{E}_t = k_B T \frac{C}{2hN} (\Delta\theta)^2 \,. \tag{2.26}$$

Besides, we measure numerically the dependence of the total energy of the system as a function of $(\Delta\theta)^2$, obtaining a relation of the type

$$E = \alpha(\Delta\theta)^2 \tag{2.27}$$

where α is a fitting parameter.

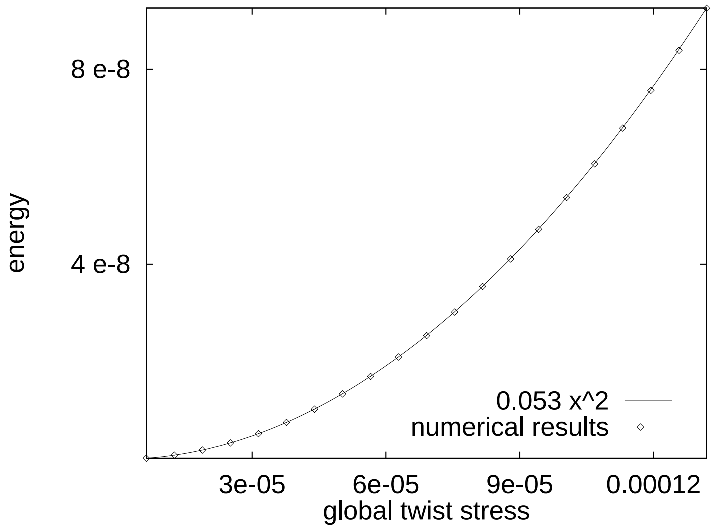


Figure 2.9: Results of energy minimization for different values of the torsional stress $\Delta\theta$ applied at the ends of a 50 bps model chain. The used minimum energy condition corresponds to the relation $\partial E/\partial s \ \leq 10^{-15}$, where s is the fictitious evolution time. The numerical data ($\diamond$) are well fitted by the curve $E = \alpha(\Delta\theta)^2$ with $\alpha = 0.053$.

Simulation results are reported in Figure 2.9. From simulations we obtain α that has then to be compared with the constant in (2.26). This gives a theoretical estimation of α which is valid if the constant radius approximation holds:

$$\alpha \stackrel{?}{=} \frac{R_0^4 \sin^2\Theta_0}{NL^2} K \approx 30\, K \tag{2.28}$$

Using in the model the constant value $K = .04\ eV\text{Å}^{-2}$ suggested from the previous theoretical derivation, the relation (2.28) gives then for α the value 0.05: the simulation gives in turns $\alpha \approx 0.05$, which confirms then that in the small twist regime the coupling with the radial degree of freedom is not relevant for the determination of the torsional rigidity of the molecule.

We can conclude that with the value $K = 0.04\ eV\text{Å}^{-2}$ our model allows to reproduce the torsional elastic properties of the real DNA molecules.

Remark 2: *The fixed planes approximation.*

Another possible problem in comparing our torsional Lagrangian term to the measured values for DNA arise from the assumption of fixed base pair planes, *i.e* of fixed h. A more realistic model should include one more degree of freedom for base pair, (and at least one more constant), *e.g.* allowing for longitudinal displacements for each base pairs. The angular degree of freedom is in fact coupled with the longitudinal one, due again to the rigidity of the backbone rods and to their geometrical configuration: stretching the molecule produces in fact a reduction of the torsional angle, as it is discussed in the many recent interpretations [52, 54, 55] of experimental results in twist-stretch manipulations on single DNA molecules [6].

The curves reported in [6] giving the relation between the longitudinal displacement dependence and the twisting number $\sigma = \Delta\theta/\Theta_0$, with a fixed stretching force applied at the two ends of the molecule, are, for small forces, symmetrical bell-like curves. This shows that the addition of some positive or negative extra twist to a chain with fixed ends produces a decrease of the chain length. With external forces of the order of $0.1\,pN$ the chain is reduced to zero relative extension at an extra twist $\sigma \approx 2\ 10^{-2}$.

Referring to these experimental results we can give an estimation of the error introduced by neglecting the longitudinal degree of freedom. According to [6] the bell shape curve is due for large twist distortions, to the formation of three-dimensional DNA loops that coils the DNA axis in a not straight configuration. Therefore, this does not correspond to local modifications of the base pair plane distance. For this reason, a direct comparison with these experimental results may not be relevant for large amplitude angular distortions. However, if the order of magnitude of a considered angular distortion is $\sigma \approx 10^{-4}$, and then corresponds to the maximum region in the twist-displacements curves reported in [6], we can infer that we are in the regime where the effect of the twist distortions considered on the longitudinal degree of freedom is quite irrelevant, so that our hypothesis of fixed base pair planes is quite good. We will test this hypothesis in the following discussions.

2.6 Dispersion relations for the model and rescaled Lagrangian

2.6.1 Rescaled approximate Lagrangian for small distortions

In the next chapter we will look for small amplitude approximate solutions of the *twist-opening* model. When dealing with small amplitude distortions, the first step is to consider the linearized system and to calculate its plane wave solutions. This corresponds to calculate the dispersion relations of the model and allows to characterize its linear stability.

The Lagrangian (2.20) can then be rewritten in terms of small displacements with respect to equilibrium. We will approximate the elastic rod Lagrangian terms up to the second order, in order to linearize the corresponding parts in the equations of motion. Let us consider the variations with respect to the equilibrium

$$\begin{aligned} Y_n &= r_n - R_0 && (2.29) \\ \Phi_n &= \varphi_n - n\Theta_0\,; && (2.30) \end{aligned}$$

for small variation $Y_n \ll 1$ and $\Phi_n \ll 1$, the potential W_{rods} in (2.20) can be written as

$$\begin{aligned} W_{rods} &\approx K \sum_n \frac{R_0^4 \sin^2\Theta_0}{L^2}(\Phi_n - \Phi_{n-1})^2 \\ &+ K \sum_n \frac{R_0^2(1-\cos\Theta_0)^2}{L^2}(Y_n + Y_{n-1})^2 \\ &+ K \sum_n \frac{2R_0^3(1-\cos\Theta_0)\sin\Theta_0}{L^2}(\Phi_n - \Phi_{n-1})(Y_n + Y_{n-1})\,. \end{aligned} \tag{2.31}$$

The Morse potential can also be approximated for small radial displacements: in this case we will keep more orders of expansion, because the non linearity arising from this term represents the main physical non linearity of the model. We will write then

$$V_m \approx D \sum \left(\alpha^2 Y_n^2 - \frac{1}{2}\alpha^3 Y_n^3 + \frac{7}{8}\alpha^4 Y_n^4\right). \tag{2.32}$$

It is useful, at this point, to rescale the system variables in order to obtain adimensional quantities. We introduce then the adimensional variations

$$y_n = \alpha Y_n \tag{2.33}$$

$$\phi_n = \alpha R_0 \Phi_n\,; \tag{2.34}$$

we rescale also the time and the model constants in an appropriate way, as defined by the following relations

$$t \;\to\; \sqrt{D\alpha^2/m}\, t \tag{2.35}$$

$$\begin{aligned} K &\to K/D\alpha^2 \\ R_0 &\to R_0\alpha \\ h &\to h\alpha \\ G_0 &\to G_0/D \end{aligned} \tag{2.36}$$

Introducing all the rescaled quantities, the Lagrangian can be in turn rescaled and it assumes the simplified form

$$\begin{aligned} \mathcal{L} \to \frac{\mathcal{L}}{D} &= \sum_n \left(\dot{y_n}^2 + \left(1 + \frac{y_n}{R_0}\right)^2 \dot{\phi_n}^2\right) - \sum_n \left(y_n^2 - \frac{1}{2}y_n^3 + \frac{7}{8}y_n^4\right) \\ &- \sum_n K_{y\phi}\big(y_n + y_{n-1}\big)\big(\phi_n - \phi_{n-1}\big) - \sum_n K_{yy}\big(y_n + y_{n-1}\big)^2 \\ &- \sum_n K_{\phi\phi}\big(\phi_n - \phi_{n-1}\big)^2 - \sum_n G\big(\phi_{n+1} + \phi_{n-1} - 2\phi_n\big)^2 \end{aligned} \tag{2.37}$$

where

$$\begin{aligned}
K_{yy} &= \left[KR_0^2/(D\alpha^2L^2)\right](1-\cos\Theta_0)^2 \,, \\
K_{\phi\phi} &= \left[KR_0^2/(D\alpha^2L^2)\right](\sin^2\Theta_0) \,, \\
K_{y\phi} &= \left[KR_0^2/(D\alpha^2L^2)\right](\sin\Theta_0)(1-\cos\Theta_0) \\
G &= G_0/(D\alpha^2R_0^2) \,.
\end{aligned} \tag{2.38}$$

Up to now we will denote with the old names the rescaled constants in Equations (2.35), (2.36), and denote with $\mathcal{L}$ the rescaled Lagrangian.

From the Lagrangian (2.37) one deduces then the approximate *equations of motion* for the system, valid in the case of small displacements:

$$\begin{aligned}
\ddot{y}_n = \big(1+\frac{y_n}{R_0}\big)\frac{1}{R_0}\dot{\phi}_n^2 - (y_n - \frac{3}{2}y_n^2 + \frac{7}{6}y_n^3) \\
- K_{yy}(y_{n+1}+y_{n-1}+2y_n) - \frac{K_{y\phi}}{2}(\phi_{n+1}-\phi_{n-1})
\end{aligned} \tag{2.39}$$

$$\begin{aligned}
\ddot{\phi}_n = -\frac{2}{R_0}\dot{y}_n\dot{\phi}_n - \frac{2}{R_0}y_n\ddot{\phi}_n - \frac{2}{R_0^2}y_n\dot{y}_n\dot{\phi}_n - \frac{1}{R_0^2}y_n^2\ddot{\phi}_n \\
+ K_{\phi\phi}(\phi_{n+1}+\phi_{n-1}-2\phi_n) + \frac{K_{y\phi}}{2}(y_{n+1}-y_{n-1}) \\
- G(\phi_{n+2}+\phi_{n-2}-4\phi_{n+1}-4\phi_{n-1}+6\phi_n)
\end{aligned} \tag{2.40}$$

In next chapters, we will refer to these equations when dealing with small amplitude distortions.

2.6.2 Dispersion relations

In order to obtain the *dispersion relations* for the model we linearize completely the equation of motion, considering just the linear terms also in the kinetic and Morse potential terms. This gives:

$$\ddot{y}_n = -y_n - K_{yy}(y_{n+1}+y_{n-1}+2y_n) - \frac{K_{y\phi}}{2}(\phi_{n+1}-\phi_{n-1}) \tag{2.41}$$

$$\begin{aligned}
\ddot{\phi}_n &= K_{\phi\phi}(\phi_{n+1}+\phi_{n-1}-2\phi_n) + \frac{K_{y\phi}}{2}(y_{n+1}-y_{n-1}) \\
&\quad -G(\phi_{n+2}+\phi_{n-2}-4\phi_{n+1}-4\phi_{n-1}+6\phi_n) \,.
\end{aligned} \tag{2.42}$$

Considering a Fourier decomposition of the lattice field (y_n, ϕ_n) into plane waves,

$$y_n = \frac{1}{2}\sum_q \hat{y}_q e^{iqn} + c.c. \tag{2.43}$$

$$\phi_n = \frac{1}{2}\sum_q \hat{\phi}_q e^{iqn} + c.c. \tag{2.44}$$

$$\tag{2.45}$$

and substituting into (2.41), (2.42), one obtains the system of equation

$$\frac{\partial^2}{\partial t^2}\begin{pmatrix}\hat{y}_q\\ \hat{\phi}_q\end{pmatrix} = \hat{J}(q)\begin{pmatrix}\hat{y}_q\\ \hat{\phi}_q\end{pmatrix} \tag{2.46}$$

$\boxed{\hat{J}(q)}$

where we have defined the matrix $\hat{J}(q)$ as

$$\hat{J}(q) = \begin{pmatrix} a(q) & c(q)\\ c^*(q) & b(q)\end{pmatrix} \doteq \begin{pmatrix} 1 + 2K_{yy}(1+\cos q) & iK_{y\phi}\sin q \\ & \\ -iK_{y\phi}\sin q & \begin{array}{c}2K_{\phi\phi}(1-\cos q)\\ +G(6 - 8\cos q + 2\cos 2q)\end{array}\end{pmatrix}. \tag{2.47}$$

The normal modes of the linearized system are calculated by finding the two eigen-values of $\hat{J}(q)$, which are the two system oscillation frequencies $\omega_\pm$. This gives, in terms of the components of $\hat{J}(q)$:

$$\omega_\pm^2(q) = \frac{1}{2}\left(a + b \pm \sqrt{(a-b)^2 + 4|c|^2}\right). \tag{2.48}$$

The system is linearly stable because the squared frequencies $\omega_\pm^2(q)$ are positive.

$\boxed{\omega_\pm}$

The corresponding eigen-vectors, $\vec{V}^\pm(q)$, give the corresponding amplitudes and phases of the oscillation in the two field components

$$\vec{V}^\pm(q) = \mathcal{N}_\pm \begin{pmatrix} 1 \\ (\omega_\pm^2 - a)/c \end{pmatrix} \tag{2.49}$$

$\boxed{\vec{V}^\pm(q)}$

where $\mathcal{N}_\pm$ are real factors introduced in order to normalize the vectors (2.49).

The dispersion curves (2.48) are shown in Figure 2.10. The parameters used in producing the plot are chosen according to the discussion Section 2.5. We report in Figure 2.10 different curves for various choices of the constant G_0.

Let us anticipate some of the topics which we will discuss in the Chapter 2, where the dispersion plays a fundamental role. The operatorial formalism that will be developed in the next chapter, in order to solve for small amplitude approximate breather solutions, is essentially an appropriate expansion of the system of equations(2.46). More exactly, one has to consider Equations (2.46) applied to a fundamental mode, so that the derivative term corresponds to multiply the eigen-vector by ω_+^2 or ω_-^2; developing then around this central mode it is possible to construct the solution as an envelope modulated linear mode, or wave packet.

As we will see, to obtain biologically meaningful solutions, it is interesting to expand around a mode with small wave number: we stress that, in the small-q range, the parameter G_0 does not really influence the dispersion curves (see Figure 2.10), and thus the model properties in the linear regime. The problem of making a good choice for this parameter will result then, fortunately, not so relevant in the context of the work presented in the following chapter.

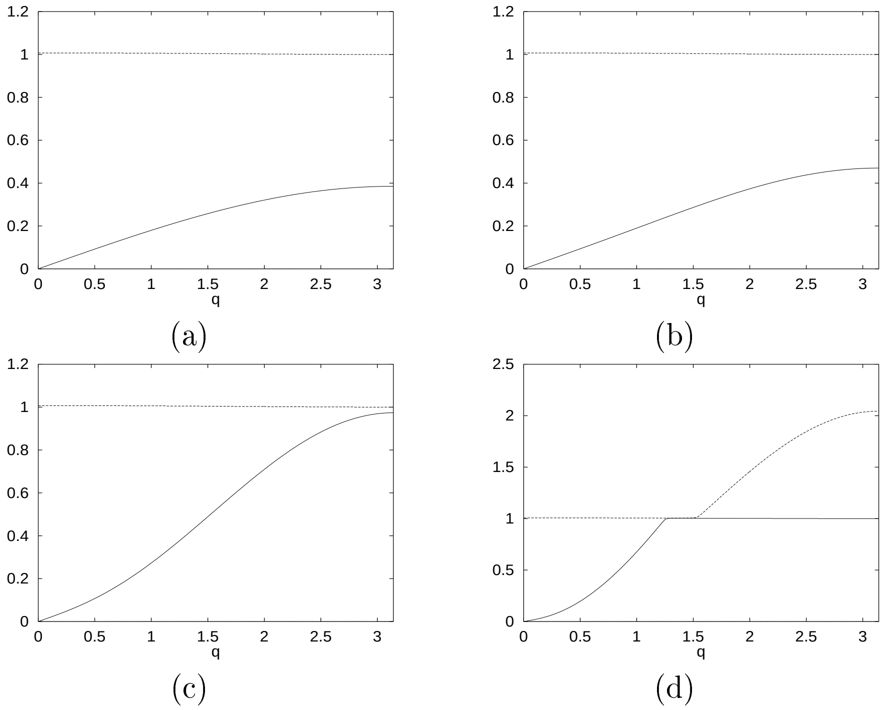


Figure 2.10: Dispersion curves for the model: q dependence of the two linear frequencies ω_+ (optical branch) and ω_- (acoustic branch) for various choices of the parameter G_0: **(a)**: $G_0 = 0.01 K R_0^2$ **(b)**: $G_0 = 0.1 K R_0^2$ **(c)**: $G_0 = K R_0^2$ **(d)**: $G_0 = 10 K R_0^2$. The other model parameters are $D = 0.04\ eV$, $\alpha = 4.45 \text{Å}^{-1}$ and $K = .04\ eV \text{Å}^{-2}$. At $q = 0$, ω_- is zero and ω_+ is $a(0) = (1 + K_{yy})^{1/2}$. At $q = \pi$, the values of the two frequencies are $b(\pi) = (4K_{\phi\phi} + 16G)^{1/2}$ and $a(\pi) = 1$: the two values corresponds to ω_- or ω_+ in dependence on the sign of $a(q) - b(q)$.

2.7 Summary

In this chapter we have first presented the PB model and its breather approximated solutions, which can be obtained by applying the MSE technique. We have discussed the interest of these kind of solutions, but we have also stessed that the PB model does not reproduce the true structure of the molecule: in order to have a realistic modeling of the transciption intiation dynamical features is instead fundamental to take into account the DNA geometry. We have then proposed a new model which reproduces this geometry by means of a couple of degrees of freedom per base pair, namely the radius and the angle of rotation. We have built the model Lagrangian by considering the various relations leading to the double helix. We have checked the model behaviour by performing simple numerical simulations. Finally, we have given the equations of motion in the case of small displacements and reported the linear dispersion relations for the model.

We will now look for breather approximate solutions of the model.