Possibility of solitary waves in the base stacks of DNA

1992 J. Phys.: Condens. Matter 4 3883

(http://iopscience.iop.org/0953-8984/4/15/004)

J. Phys.: Condens. Matter 4 (1992) 3883–3903. Printed in the UK

Possibility of solitary waves in the base stacks of DNA

Detlef Hofmann, Janos Ladik, Wolfgang Förner and Peter Otto
Chair for Theoretical Chemistry and Laboratory of the National Foundation for Cancer Research at the Friedrich-Alexander-University Erlangen-Nürnberg, Egerlandstrasse 3, D-8520 Erlangen, Federal Republic of Germany

Received 30 July 1991, in final form 14 January 1992

Abstract. A realistic model for the dynamics of B-DNA is developed and presented. Within this model we study the time evolution of conformational excitations of single nucleotide bases and base pairs in simple base stacks and in DNA helices. Further the possibility of the existence of solitary waves in DNA is investigated. On the basis of our classical model we do not find such a wave in DNA, but for a stack of adenine molecules without a backbone we predict one. Possible extensions of the model for DNA are discussed. From our results we can conclude that solitons exist in stacked systems without an additional backbone.

1. Introduction

With the development of large computers, non-linear phenomena, described by non-linear differential equations without an analytical solution, have become more and more of interest. One of the most interesting results from such equations is the existence of solitons. In a mathematically strict sense solitons are analytical solutions of special non-linear differential equations. Solitons are non-dispersing wave packets, which are able to cross one another without perturbation. They do not change their shape and they do not lose their energy in time. Solitons are examined mathematically for instance in [1]. In physics solitons are used to describe ferromagnetism [2], antiferromagnetism [3], phase changes [4], charge transport in *trans*-polyacetylene (PA) [5], dynamics of the sugar–phosphate backbone of DNA [6], and dynamics of the two coupled base stacks of the DNA double helix bound by hydrogen bridges [7]. The first observation of such a solitary wave happened back in 1844 [8] in the form of a water wave in a channel. In optics the transport of light pulses through special fibres is soliton-like [9] and solitons are observed also in Josephson junctions as fluxons [10]. In the case of PA it could be shown that the numerical solutions for the continuum limit are identical to the exact mathematical ones [11]. Another possible application is the transport of conformational energy in stacked systems. One of the best-known systems of this kind is the base stacks in DNA. This could provide a possible mechanism for carcinogenesis [12, 13], since the solitons can interfere with DNA–protein interactions and thus might be able to remove blocking proteins from oncogens. Owing to their solitary nature, this interference can occur from the site at which the solitons were initially excited by the conformational change caused by carcinogen binding and subsequent release. Such long-range effects of carcinogens in DNA are necessary to explain their role in carcinogenesis, as was shown

0953-8984/92/153883 + 21 $04.50

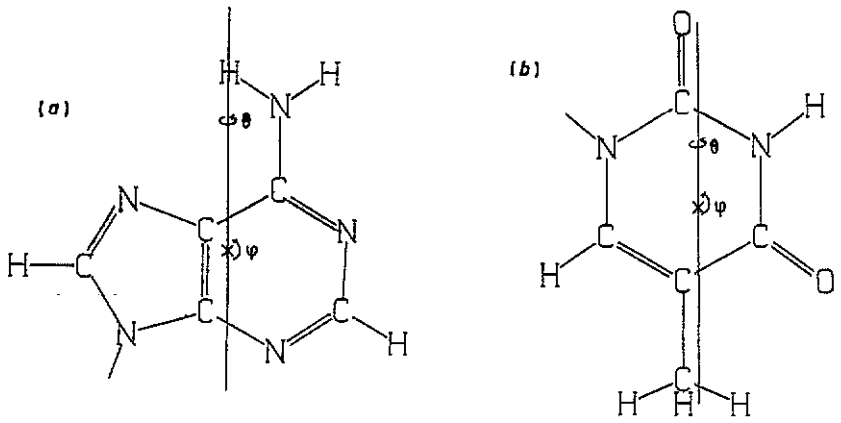


Figure 1. Definition of the rotation axes in (*a*) adenine and (*b*) thymine.

by simple statistical considerations, which exclude the possibility of a purely local effect of the chemical carcinogens [12].

Since we are considering numerical solutions of systems of differential equations, it is not possible to prove that the definition for a soliton is fulfilled exactly, and thus it is more correct to speak of solitary waves. In preliminary model calculations we have studied a stack of formamide molecules [14] and the influence of impurities in the form of thioformamide molecules [15]. The results confirmed the existence of solitary waves in these model systems. A comparison of our model, where the molecules are treated as classical particles moving on an analytically given potential surface, with adiabatic ones in which the π-electrons were treated explicitly on the Hückel and Pariser–Parr–Pople (PPP) level justified the use of our classical *ansatz* for the potential energy surface [16]. However, the parameters of the surface for the base–base interactions are calculated using methods from *ab initio* quantum chemistry.

2. Method

The *ansatz* for the total energy of our system is

$$E_{\text{tot}} = E_{\text{kin}} + E_{\text{pi}} + E_{\text{vi}} + E_{\text{bb}}. \tag{1}$$

The kinetic energy is given by E_{kin}, and the planar interaction between two base pairs by E_{pi}. E_{pi} is mainly determined by the two (three for the guanine and cytosine base pair) hydrogen bonds of adenine and thymine. The vertical stacking interaction between two bases is given by E_{vi}, and the interaction of the sugar–phosphate backbone with a base by E_{bb}. Each base has three degrees of freedom, movement along the axis of the helix (z axis) and two rotations. One is a rotation denoted by φ in the plane perpendicular to the helix axis and the other one ϑ is the tilting angle of the base (figure 1).

Because the bases are not treated as point masses but as rigid bodies, the expression for the kinetic energy is [17]

$$T_{\text{n}} = \tfrac{1}{2} q_n^{+} \mathbf{K}_n q_n \tag{2a}$$

with

$$K_n(1,1) = M \qquad K_n(2,1) = M_z \sin\vartheta_n + M_x \cos\vartheta_n \qquad K_n(2,2) = M_{xx} + M_{zz}$$
$$K_n(3,1) = 0 \qquad K_n(3,2) = M_{yz}\cos\vartheta_n - M_{xy}\sin\vartheta_n \tag{2b}$$
$$K_n(3,3) = M_{xx}\cos^2\vartheta_n + M_{yy} + M_{zz}\sin^2\vartheta_n + 2M_{xz}\sin\vartheta_n\cos\vartheta_n.$$

Also a detailed derivation of the equations of motion can be found in [18]. The abbreviations are the following:

$$M = \sum_i m_i \qquad M_\mu = \sum_i m_i\mu_{i0} \qquad M_{\mu\nu} = \sum_i m_i\mu_{i0}\nu_{i0} \qquad q_n^+ = (z_n, \vartheta_n, \varphi_n). \tag{3}$$

Here μ and ν are the space coordinates x, y and z; μ_{i0} is the coordinate μ ($\mu = x, y, z$) of atom i in the base at site 0 in the equilibrium position. The equilibrium positions follow from the geometry of B-DNA.

All energies of interaction were expanded with the help of a five-dimensional Taylor series in the degrees of freedom. Because the rotation axes for two interacting bases do not coincide, we have to distinguish between rotation of base 1 and base 2 (denoted by α and β), which both have to be treated explicitly. In contrast the movement in the z direction appears only in the difference $\Delta z = z_1 - z_2$. Therefore the Taylor series is written in general ($\boldsymbol{k}$ is a vector containing the exponents, w contains the independent degrees of freedom) for the potential energies E_{vi}, E_{pi} and E_{bb}:

$$E = \sum_{k_i \geq 0}^{\sum_{i=1}^{5} k_i \leq \sigma} D_k \prod_{i=1}^{5} w_i^{k_i} \tag{4}$$
$$w^+ = (\Delta z, \vartheta_1, \vartheta_2, \varphi_1, \varphi_2).$$

Here σ is a constant, which can be freely chosen. In all cases we have a dependence on site n. For the planar interactions $E_{\mathrm{pi},n}$ we have only two different types t' of potential surfaces, namely adenine–thymine or guanine–cytosine. To consider this, we write

$$D_k \rightarrow D_{\mathrm{pi},k,n,t'} \qquad w^+ \rightarrow (z_{n\alpha} - z_{n\beta}, \vartheta_{n\alpha}, \vartheta_{n\beta}, \varphi_{n\alpha}, \varphi_{n\beta}). \tag{5}$$

Here α and β define the two different strands of DNA. With the help of symmetry arguments, one can see that for the pairs the other way round the same series is valid, e.g. one can use the same parameters for adenine–thymine as for thymine–adenine. Unfortunately this is not the case for E_{vi}. Here 16 different potential surfaces exist for the 4×4 possible combinations of bases t_n:

$$D_k \rightarrow D_{\mathrm{vi},k,n,t_n t_{n+1}}$$
$$w^+ \rightarrow (z_{n+1,\gamma} - z_{n,\gamma}; \vartheta_{n+1,\gamma}; \vartheta_{n,\gamma}; \varphi_{n+1,\gamma}; \varphi_{n,\gamma}). \tag{6}$$

Here γ denotes the strand α or β. There are four possibilities for the interaction between sugar s and a base t:

$$D_k \rightarrow D_{\mathrm{bb},k,n,t} \qquad w^+ \rightarrow (z_{n\gamma} - z_{s\gamma}, \vartheta_{n\gamma}, \vartheta_{s\gamma}, \varphi_{n\gamma}, \varphi_{s\gamma}). \tag{7}$$

In our model the degrees of freedom ($z_{s\gamma}$, $\vartheta_{s\gamma}$, $\varphi_{s\gamma}$) for sugar are kept fixed and are not treated explicitly up to now. For a unique description of all interactions, it was preferred to write the potential surfaces in this way. Also this will allow one to include the dynamics of the sugar–phosphate backbone in future without change in the formalism.

We used two methods to obtain the necessary parameters D since direct *ab initio* calculations would be too time-consuming owing to the size of the molecules involved

and the large number of points that have to be computed on the potential surfaces. For the covalent bonds of the bases to the backbone, an all-atom force-field model was used [19]. This method divides the energy into a sum of terms for bond lengths, angles between bonded atoms, torsional angles, hydrogen bonds, electrostatic interactions and non-bonding interactions (van der Waals interactions):

$$E_{\text{tot}} = \sum_{\text{bonds}} K_R(R - R_0)^2 + \sum_{\substack{\text{bond}\\ \text{angles}}} K_\vartheta(\vartheta - \vartheta_0)^2 + \sum_{\substack{\text{dihedral}\\ \text{angles}}} \sum_n \frac{V_n}{2}[1 + \cos(n\varphi - \gamma)]$$
$$+ \sum_{\substack{\text{hydrogen}\\ \text{bridges}}} \left(\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right) + \sum_{\text{atoms}i<j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{R_{ij}}\right). \tag{8}$$

Here R_0 is the standard bond length and ϑ_0 the standard bond angle. The term for the dihedral angles is a Fourier expansion; however, as usual we use only one term of the series, where it depends on the actual dihedral angle which one is applied (number of minima in the potential for the rotation). The necessary parameters were in general determined from infrared and Raman spectra and a set optimized for DNA was taken from [19]. The atomic changes q_i we have calculated with the help of *ab initio* Hartree–Fock calculations using an atomic basis set of double-ζ quality [20]. Since we consider only small displacements, this *ansatz* can be used for our purpose. Necessarily an *ansatz* like (8) has its limitations. First of all the asymptotic behaviour for large bond lengths R is wrong. If larger displacements are considered, one should replace the Taylor series in $(R - R_0)$ by an $(R - R_0)^{-n}$ expansion, although this would require a larger number of parameters and consequently would increase the necessary computation time. Owing to the linear *ansatz* for the forces in (8), also overtones in the vibrational spectrum are described incorrectly. Furthermore, also the bond angles would be better described via an expansion into a Fourier series instead of the applied Taylor series. However, for small displacements as considered by us, *ansatz* (8) is sufficient. The non-binding parameters ε and r^* given in [19] are related to the parameters C and D in the Lennard-Jones 10–12 potential for hydrogen bridges via

$$C = \tfrac{1}{2}\varepsilon(r^*)^{12} \qquad D = 2\varepsilon(r^*)^6 \tag{9}$$

where ε is the energy of a hydrogen bridge and r^* is the displacement relative to one of its minima. However, we compute only the sugar–base interaction with this *ansatz* and therefore no hydrogen bridges occur in our case. The geometry of the sugar fragment used in our calculations is shown in figure 2.

The potential surfaces between stacked or hydrogen-bonded bases were calculated with the pseudo-polarization tensor mutually consistent field (PPT-MCF) method [21–24], which results in interaction energies of *ab initio* Hartree–Fock (double-ζ basis set) quality [24]. This method is of higher quality than the force field described above, but can only be used for non-bonded interacting molecules. Therefore it is not possible to apply it to the sugar–base interactions, which include a covalent bond (C–N) between the interacting fragments. In the PPT-MCF method the electrostatic potential of a molecule I calculated on the *ab initio* Hartree–Fock level is fitted by a number (N_{PC}^I) of point charges q_i^I at points r_i^I. The position vectors of these point charges of the free molecule are shifted owing to the electric field of another molecule with which it interacts. This shift is obtained with the help of pseudo-polarization tensors. These tensors are calculated by computing the point-charge distributions of the free molecule in electric

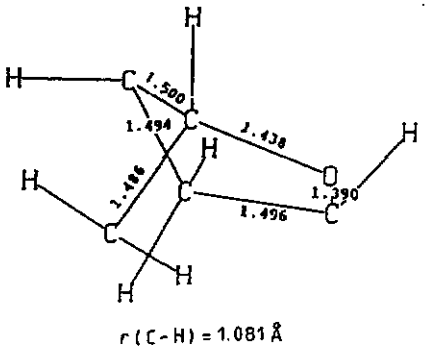


Figure 2. Geometry of the sugar fragment as used in our calculations.

fields of different strengths. Having once fixed the point-charge distribution and the PPT of a molecule, one can apply these for calculation of interaction energies with any other molecule. Since the point-charge distributions in this way are allowed to relax upon interactions, the polarization part of the interaction energy is directly included in the electrostatic interaction energy between the point-charge distributions.

The total interaction energy between two molecules I and J is then computed as

$$\Delta E^{IJ} = E^{IJ}_{\text{elst}} + E^{IJ}_{\text{exch}} + E^{IJ}_{\text{disp}} \tag{10}$$

where E_{elst} is the electrostatic interaction energy plus the polarization energy, E_{exch} the intermolecular exchange energy and E_{disp} the dispersion energy. E_{elst} is given by

$$E^{IJ}_{\text{elst}} = \sum_i^{N^I_{\text{PC}}} \sum_j^{N^J_{\text{PC}}} \frac{q^I_i q^J_j}{|\boldsymbol{r}^I_i - \boldsymbol{r}^J_j|} - \sum_i^{N^I_{\text{PC}}} \sum_l^{M^J} \frac{q^I_i Q^J_l}{|\boldsymbol{r}^I_i - \boldsymbol{R}^J_l|} - \sum_k^{M^I} \sum_j^{N^J_{\text{PC}}} \frac{Q^I_k q^J_j}{|\boldsymbol{R}^I_k - \boldsymbol{r}^J_j|} + \sum_k^{M^I} \sum_l^{M^J} \frac{Q^I_k Q^J_l}{|\boldsymbol{R}^I_k - \boldsymbol{R}^J_l|}. \tag{11}$$

In (11) the nuclear charges are Q^I_i, $\boldsymbol{R}^I_k$ is the position vector of the kth nucleus in the Ith molecule and M^I is the number of nuclei in molecule I. In an actual calculation one computes the electrostatic potential of the charges of molecule J at the positions of all point charges of molecule I. Then with the help of the pseudo-polarization tensors the shifts of the point charges of molecule I due to the presence of the charges of molecule J are calculated. Then the corresponding shifts in J due to the presence of I are determined. With the new point-charge distributions of I and J the procedure is repeated until the iteration converges. In this way the polarization energy is included in E_{elst}.

The exchange energy is given approximately by

$$E^{IJ}_{\text{exch}} = -\sum_i^{N^I_{\text{PC}}} \left(\frac{81}{\pi}\rho^J(\boldsymbol{r}^I_i)\right)^{1/3} \rho^I(\boldsymbol{r}^I_i) w_i - \sum_j^{N^J_{\text{PC}}} \left(\frac{81}{\pi}\rho^I(\boldsymbol{r}^J_j)\right)^{1/3} \rho^J(\boldsymbol{r}^J_j) w_j. \tag{12}$$

The necessary electron density ρ is computed with the help of the Hartree–Fock wavefunction of the free molecules. The weighting factor w_i is given by q^I_i/N^I_e where N^i_e is the number of electrons in molecule I.

Finally, the intermolecular dispersion energy is taken into account employing a London-type expression, together with empirical parameters [18]:

$$E_{\text{disp}} = -\frac{3}{2} \sum_i^{M^I} \sum_j^{M^J} \frac{\alpha_i \alpha_j \text{IP}_i \text{IP}_j}{(\text{IP}_i + \text{IP}_j) R^6_{ij}} \tag{13}$$

Table 1. Comparison of the largest error ΔE_{max} and the average error ΔE_{ave} in the fit of our potential surfaces by Taylor series with the largest (E_{max}) and smallest (E_{min}) interaction energy occurring (all energies in millihartrees) together with the number of points m computed on each individual surface (index v indicates a stacked base pair, h is a hydrogen-bonded one; S indicates sugar, A adenine and T thymine).

Surface	ΔE_{max}	ΔE_{ave}	E_{max}	E_{min}	m
$A–A_v$	0.985	0.054	6.434	−6.879	911
$A–T_v$	0.446	0.024	8.190	−2.978	997
$T–T_v$	2.546	0.148	18.305	−31.554	997
$T–A_v$	9.998	0.214	9.445	−18.727	988
$A–T_h$	0.403	0.021	3.143	−10.305	969
S–A	483.192	2.077	1258.513	−13.227	997
S–T	163.370	3.489	924.999	−6.298	997

where α_i is the polarizability and IP_i the ionization potential of atom i (in molecule I) in its valence state. R_{ij} is the distance between the nuclei i in I and j in J. To get an idea of the quality of this approximation, we computed the dispersion energy between two stacked adenine molecules in B-DNA geometry and obtained with the help of (13) a value of −0.2891 eV [25]. This is in very good agreement with the *ab initio* (Hartree–Fock valence split basis set calculation and perturbation theory for interacting molecules including overlap for the dispersion term) result of −0.2865 eV [26].

For all potential surfaces necessary, m points on each of them were calculated and the so-called l_1-norm $r(w)$ was minimized, to fit our Taylor series *ansatz* to the calculated points:

$$r(w) = \sum_{l=1}^{m} |r_l| \qquad \text{with} \qquad r_l = E_l - \sum_{k_i \geq 0}^{\sum_{i=1}^{5} k_i \leq \sigma} D_k \prod_{i=1}^{5} w_l^{k_i}. \tag{14}$$

With this method the sum of distances of the points calculated by the Taylor series and by the *ab initio* PPT-MCF method at randomly chosen points l on the potential surface is minimized. The number of parameters is for an n-dimensional Taylor series ($n = 5$) of order σ ($\sigma = 6$):

$$\binom{\sigma + n}{n} = 462. \tag{15}$$

Table 1 shows the maximal interaction energy E_{max} in the fitted range, the minimal interaction energy E_{min}, the maximal error, the average error and the number of points m. ΔE_{max} is quite large using this norm, but these errors are near to the borders of the fit region for the potential surfaces and for that reason not very important. This region is never reached in a dynamic simulation. The fitted range was for the angles ±0.2 rad and for the displacement in the z direction ±1.2 Bohr. The relative average error is very small (about 0.5%), especially near to equilibrium, and therefore this method was preferred against a least-squares or l_∞-norm. If the interactions are of vertical type, $A–B_v$ denotes B following A in the z direction. The horizontal interaction $A–B_h$ denotes a hydrogen-bonded base pair with A in the α and B in the β strand. The structures of the units were implemented in the PPT-MCF program [27] and were taken from the crystal structure of B-DNA [28, 29].

After having obtained all the necessary potential surfaces in analytical form, the equations of motion of the system were numerically solved with the help of a simple one-step procedure. These equations can be obtained from the expressions for the kinetic (T) and potential energy (V) via Euler–Lagrange equations of the second kind (see [17] for details). They read as

$$\mathbf{K}_n \ddot{\boldsymbol{q}}_n + \boldsymbol{Q}_n = \boldsymbol{F}_n \tag{16}$$

where $\boldsymbol{q}_n$ denotes the vector of the three geometrical degrees of freedom of each base n (z_n, ϑ_n, φ_n) and the forces can be computed from the potential energy as

$$\boldsymbol{F}_n = -\partial V / \partial \boldsymbol{q}_n. \tag{17}$$

Further

$$\begin{aligned} Q_n(1) &= (M_z \cos\vartheta_n - M_x \sin\vartheta_n)\dot{\vartheta}_n^2 \qquad Q_n(2) = Y_n \dot{\varphi}_n^2 \\ Q_n(3) &= -(M_{xy} \cos\vartheta_n + M_{yz} \sin\vartheta_n)\dot{\vartheta}_n^2 - 2Y_n \dot{\vartheta}_n \dot{\varphi}_n \\ Y_n &= (1 - \cos^2\vartheta_n)M_{xz} + (M_{xx} - M_{zz}) \sin\vartheta_n \cos\vartheta_n. \end{aligned} \tag{18}$$

Since in our case $\mathbf{K}_n(\vartheta_n)$ is non-singular over the whole range of the variable, this can be rewritten as

$$\ddot{\boldsymbol{q}}_n = \mathbf{K}_n^{-1}(\boldsymbol{F}_n - \boldsymbol{Q}_n). \tag{19}$$

Then from the geometry and the velocities at a given time step, one can compute the accelerations and, assuming these accelerations to be constant over the time step length τ, the geometries and velocities at the next time step can be calculated. In this way after the time $m\tau$ we obtain

$$\begin{aligned} \dot{\boldsymbol{q}}_n[(m+1)\tau] &= \dot{\boldsymbol{q}}_n[m\tau] + \tau \mathbf{K}_n^{-1}(\boldsymbol{F}_n[m\tau] - \boldsymbol{Q}_n[m\tau]) \\ \boldsymbol{q}_n[(m+1)\tau] &= \boldsymbol{q}_n[m\tau] + \tau \dot{\boldsymbol{q}}_n[(m+1)\tau]. \end{aligned} \tag{20}$$

A time step length of 2 fs (femtoseconds) is sufficiently small to keep the error in total energy less than 0.1% of the kinetic energy. Further reduction of the time step size did not change the results. For the simulations performed in this work, the one-step procedure is sufficient; however, in future simulations on longer chains with larger simulation times one has to introduce a higher-order procedure (e.g. a Runge–Kutta or a predictor–corrector method) in order to reduce the computation time.

3. Results

First of all we computed 911 randomly chosen points on the potential surface of a stacked adenine dimer (A–A_v) in the above-mentioned region of ± 0.2 rad for the angles and ± 1.2 Bohr for the displacement z along the helix axis around the equilibrium geometry of B-DNA (3.36 Å distance and rotation of 36° around φ between the two bases). After the fitting procedure described above we computed a number of additional points on the surface with the PPT-MCF method and compared the results with values calculated from our analytical potential surface in order to check the quality of our fit. The results are displayed in figure 3. In each of the curves only one of the variables was varied, while the others were kept at their equilibrium values.

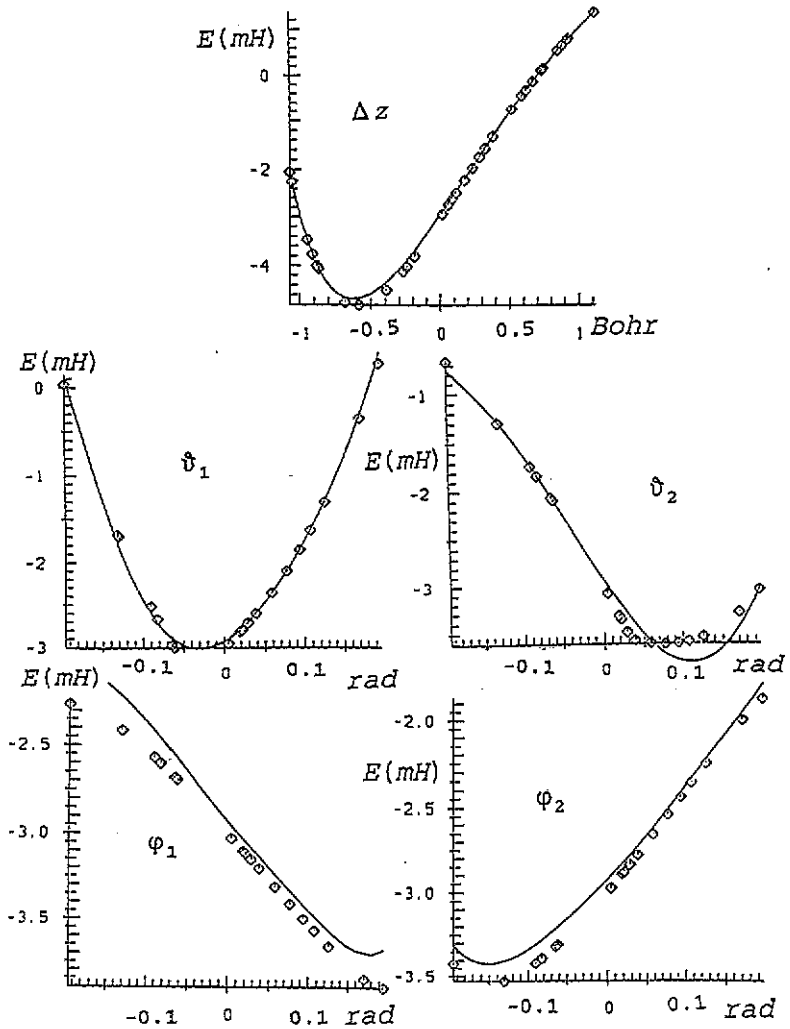


Figure 3. Comparison of potential curves from our fit (full curves) and points computed directly with the PPT-MCF method for the five coordinates for A–A_v (energies in millihartrees).

Obviously the agreement between the directly calculated points and the curves obtained from the fit is satisfactory. Further we note that the potential as a function of φ has no minimum in the fit region. Therefore, we decided to keep the angles φ and ϑ frozen in our first dynamic simulations of single- and double-stranded nucleotide base stacks (without inclusion of the backbone potential). This seems to be justified because in our previous simulations [15] on formamide stacks we observed that the angle φ had a rather unimportant influence on the dynamics, while excitations of the angle ϑ behaved similar to excitations in the displacement z [15]. In these simulations of pure adenine and thymine stacks (single strand, no backbone) we used open chain ends and initial displacements in z of the second unit. In figure 4 we show the time evolution of the local kinetic energy T_n of an adenine stack for different initial excitations as a function of site n.

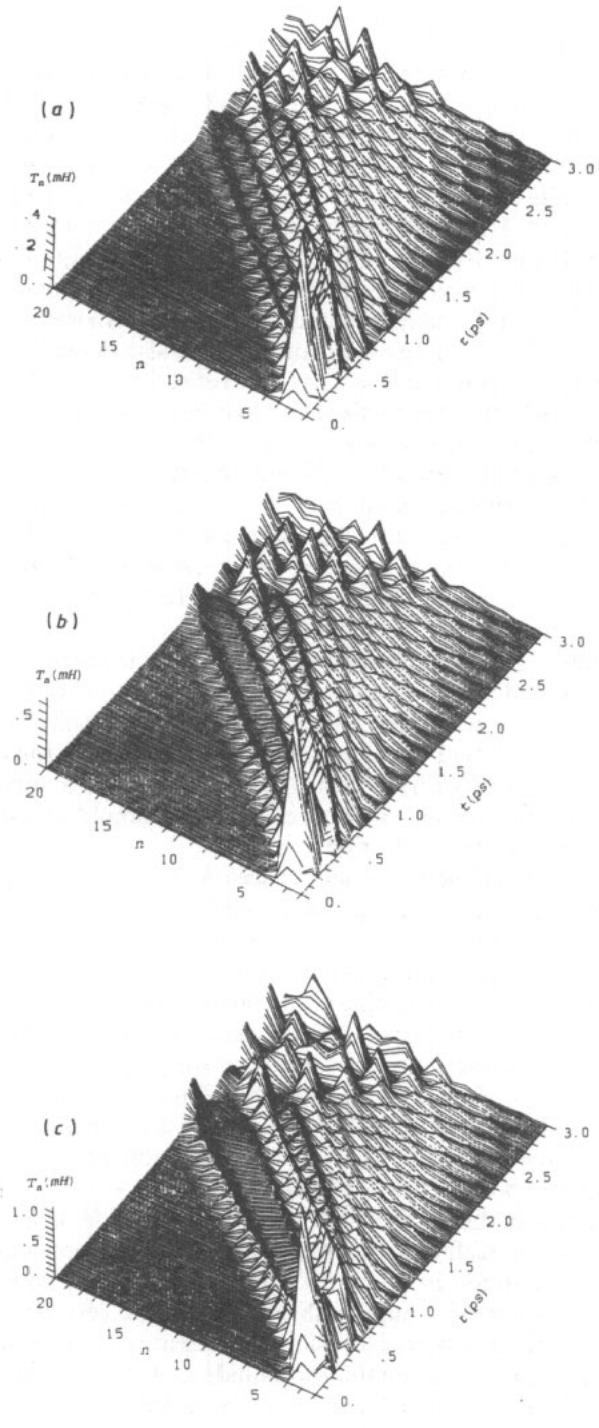
Initial excitations of the second site between −0.1 and −0.3 Bohr did not lead to formation of solitary waves. As an example of this behaviour the case $z(2) = -0.3$ Bohr is shown in figure 4(*a*). The excitation clearly disperses rapidly over the entire chain. In the case of $z(2) = -0.4$ Bohr (figure 4(*b*)) we still observe considerable dispersion; however, also a solitary wave starts to form, which can be clearly seen in figure 4(*c*) with $z(2) = -0.5$ Bohr. In this case even two waves of solitary character are emitted from the initial excitation, one immediately, the second one with a considerable delay. Both waves have similar amplitudes and velocities and they seem to be unable to survive the reflection at the chain end. However, the latter observation is not clear owing to the short simulation time. Figure 4(*d*) shows the case of an excitation (−0.4 Bohr) in the middle of the chain. Now two waves are emitted in opposite directions, and obviously they do survive reflection at the chain ends and the collision with each other, although after reflection the dispersion of energy is enhanced. Thus also in the former case of the $z(2)$ excitation one expects that the waves survive the reflection at the chain end. Finally figure 4(*e*) shows a case where we have fixed the two terminal molecules. In this case already for $z(2) = -0.2$ Bohr a solitary wave forms with an estimated effective mass of $740m_e$ and a velocity of about 1.6 km s^{-1}. However, it seems to us that open chain ends should be a more realistic boundary condition. The solitary waves observed are quite similar to those in formamide stacks [15] where also a lower threshold for solitary wave formation was found to exist in the excitation energy. However, the formation of two waves from a local excitation was not found in that case. In figure 5 we show the results of similar simulations on pure thymine stacks.

We observe the same behaviour as in the case of adenine with solitary wave formation starting from $z(2) = -0.4$ Bohr. Figure 4(*b*) shows that in the thymine case the two waves form with less dispersion of energy than in the adenine stack. Also the solitary waves in the thymine stack appear to be considerably faster than in the adenine stack. The long-time simulation (figure 4(*c*)) indicates that the second wave survives reflection at the chain end, while the first one, which has a smaller amplitude, disperses upon collision with the second one after reflection.

In a next step we calculated the potential surfaces $A–T_v$ and $T–A_v$. With the help of these surfaces we could test the influence of aperiodicity on the solitary waves. Before starting a simulation it was necessary in this case to perform a geometry optimization, because the thymine molecule in the middle of the adenine stack causes a geometry distortion. This can be done by performing a simulation on the system, but multiplying the velocities of the bases with a factor of 0.95 after each time step. In this way one does not obtain a dynamical simulation, but a smoothly converging geometry optimization, which is complete when the kinetic energy vanishes. This procedure is identical to a modified gradient optimization method. Figure 6 shows the influence of one thymine molecule at site 10 in an adenine stack, where in this simulation the chain ends were kept fixed. The initial excitation was $z(2) = -0.2$ Bohr as in figure 4(*e*).

As we have observed earlier [15] in the case of formamide stacks disturbed by thioformamide molecules, the influence of impurities on solitons depends on differences in the masses and on the different electronic interactions of the impurity molecule compared to the host system. Because both influences are very large for thymine molecules in an adenine stack, the transmission coefficient is even smaller than in the case of formamide stacks perturbed by thioformamide molecules. This should change for stacks of adenine–thymine base pairs disturbed by guanine–cytosine pairs because the masses and the structures are more similar.

In further computer experiments we have studied double-stranded stacks of hydro-


(a)
T_n (mH)
.4
2
0.
20
15
n
10
5
0.
.5
1.0
1.5
2.0
2.5
3.0
t (ps)
(b)
T_n (mH)
.5
0.
20
15
n
10
5
0.
.5
1.0
1.5
2.0
2.5
3.0
t (ps)
(c)
T_n (mH)
1.0
.5
0.
20
15
n
10
5
0.
.5
1.0
1.5
2.0
2.5
3.0
t (ps)

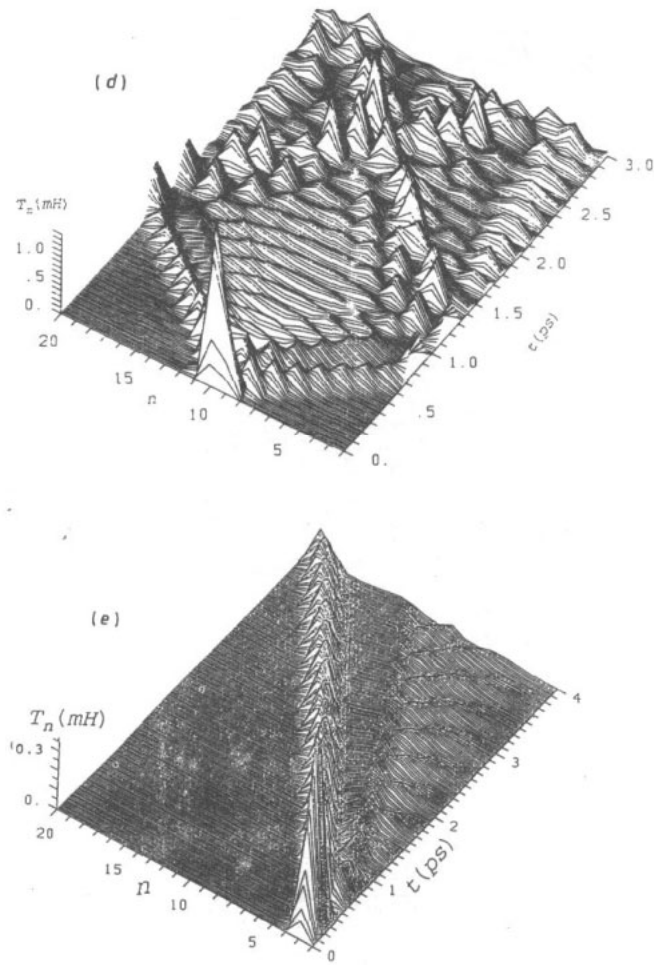


Figure 4. The local kinetic energy T_n (mHartree) as a function of time and site n in a stack of adenine molecules (single strand without backbone, ϑ and φ frozen; with the exception of (*e*) the chain ends were allowed to move freely) for different initial excitations: (*a*) $z(2) = -0.3$ Bohr; (*b*) $z(2) = -0.4$ Bohr; (*c*) $z(2) = -0.5$ Bohr; (*d*) $z(10) = -0.4$ Bohr; (*e*) $z(2) = -0.2$ Bohr, fixed chain ends.

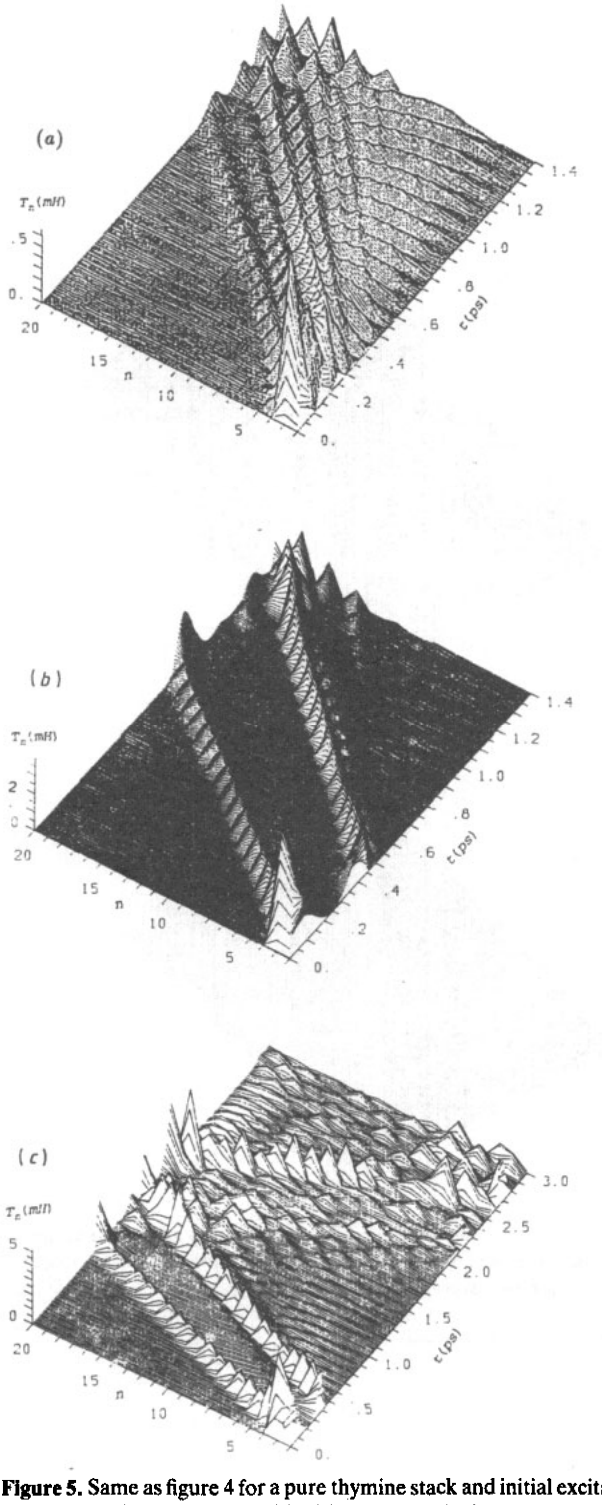


Figure 5. Same as figure 4 for a pure thymine stack and initial excitations of: (*a*) $z(2) = -0.2$ Bohr; (*b*) $z(2) = -0.4$ Bohr; (*c*) $z(2) = -0.4$ Bohr (larger simulation time).

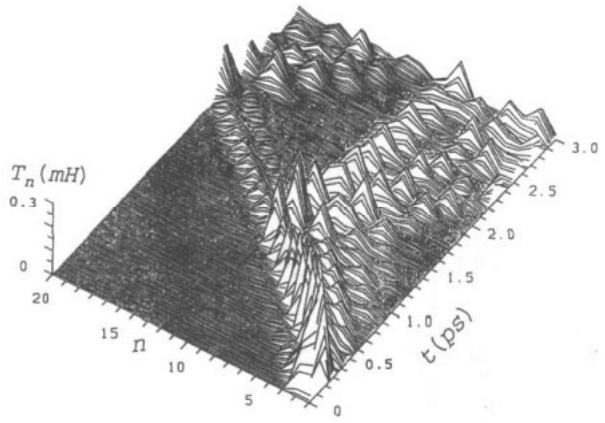


Figure 6. Same as figure 4 for a stack of adenine molecules with one thymine unit at site 10; fixed chain ends, geometry optimization prior to simulation and initial excitation of $z(2) = -0.2$ Bohr.

gen-bonded adenine–thymine (A–T) base pairs. Also in these calculations we did not include the backbone potential and worked with open chain ends. The angle ϑ was kept frozen again. In the simulations shown in figures 7(*a*)–(*c*) φ was frozen also, while in figures 7(*d*)–(*f*) it was allowed to be free. We see that in the double-stranded case for a small excitation in $z(2)$ no solitary wave is formed. In figure 7(*b*) after a delay of about 0.5 ps a solitary wave is emitted from the initial excitation of $z(2) = -0.4$ Bohr. After this the rest of the excitation disperses. Immediately after the start of the simulation a wave of rather small amplitude also starts to propagate. A similar behaviour is observed for the larger excitation in figure 7(*c*). Although the movement of the solitary wave is not as regular as for the smaller excitation in this case, the energy is here more concentrated in the wave than before and the energy dispersion is reduced. In figures 7(*d*) and (*e*) we observe that the behaviour of the system is not changed very much when the angles φ are allowed to change. Figure 7(*f*) shows a solitary wave emitted from an excitation in both variables ($z(2) = -0.4$ Bohr, $\varphi(2) = -0.02$ rad $= -1.1°$). For larger excitations in φ the stack turned out to be unstable owing to the lack of a backbone potential. Further, in figure 8(*a*) we show the case of a double-stranded A–T stack with a reversed T–A pair in position 9. The initial excitation was $z(2) = -0.4$ Bohr, and the angles were frozen again.

The transmission coefficient for the wave through the perturbation is rather small and most of the energy is reflected. Thus it is not surprising that in an alternating stack of A–T and T–A pairs (figure 8(*b*)) no solitary wave is formed. The situation is similar to the case of a single-stranded A stack with a T impurity discussed above. The only difference is that here the hydrogen-bonded pairs are always moving together. Thus there is no difference in the masses of the moving entities (hydrogen-bonded A–T and T–A pairs). Therefore we have to conclude that the differences in the interactions between stacked AA (TT) and AT pairs constitute a large enough disorder to reflect most of the energy of the solitary waves. Thus we can conclude that starting from initial excitations as considered by us and in the absence of disorder solitary waves exist in stacked single and double helices of nucleotide bases, provided that the initial excitation is large enough and that no backbone is present.

To improve the model for DNA further, in a next step the interaction of the sugar

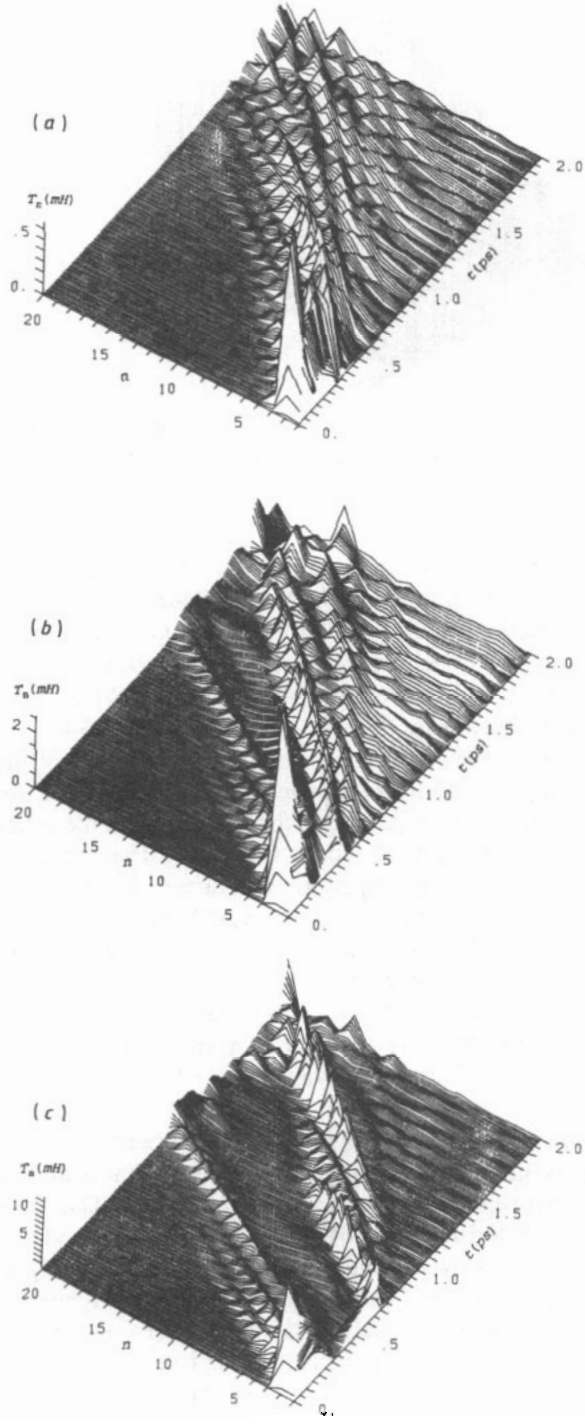


Figure 7. Time evolution of the local kinetic energy T_n as a function of site n for a double-stranded A–T stack (open chain ends, no backbone potential). The figures show only one of the two strands, because the picture for the second one was identical in each case. In (*a*)–(*c*) the angles φ were kept fixed;

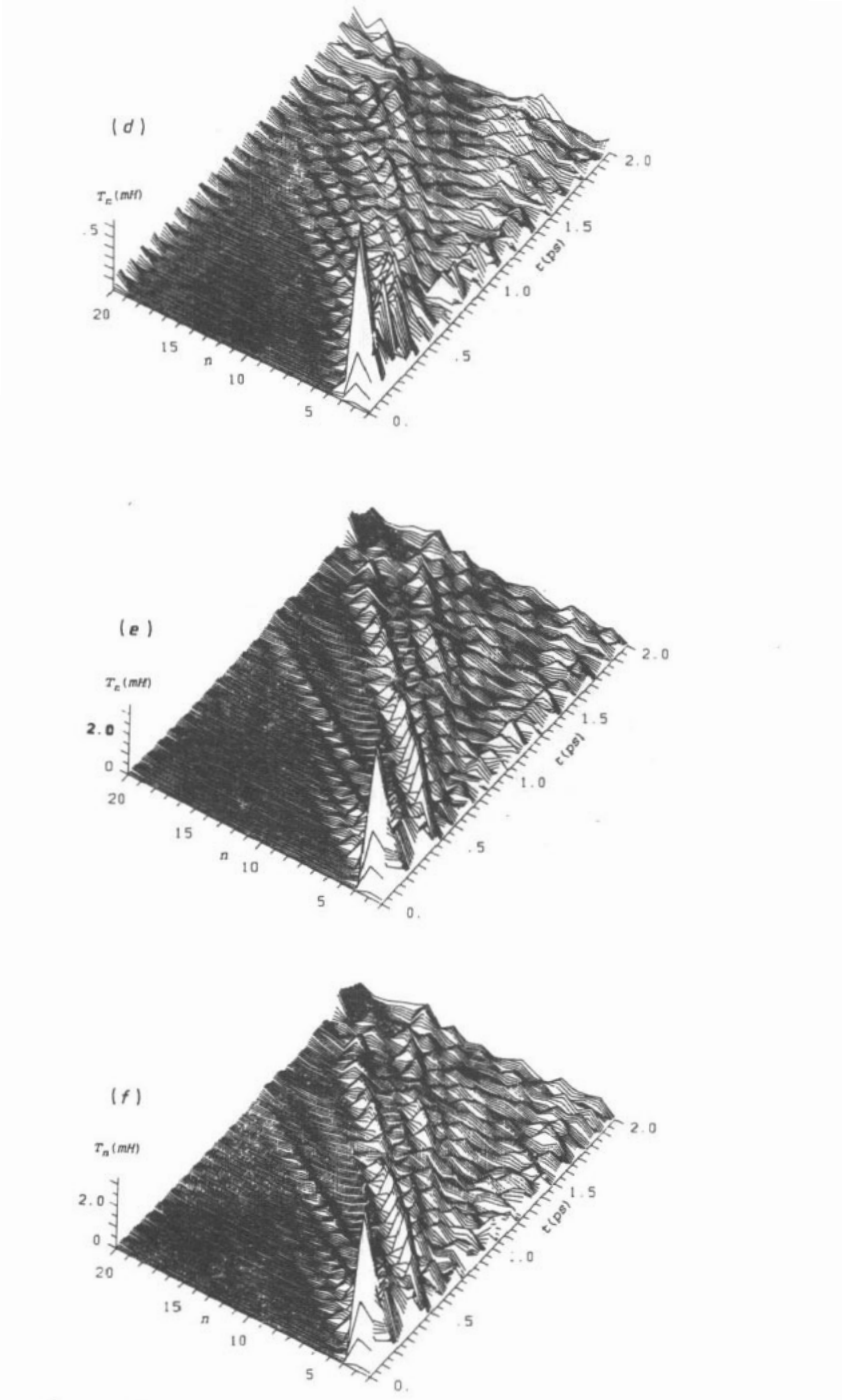


Figure 7 *continued*: in (*d*)–(*f*) they were allowed to move. The angles ϑ were frozen in all simulations. The initial excitations (always a complete hydrogen-bonded A–T pair was displaced) are: (*a*) $z(2) = -0.2$ Bohr; (*b*) $z(2) = -0.4$ Bohr; (*c*) $z(2) = -0.6$ Bohr; (*d*) $z(2) = -0.2$ Bohr; (*e*) $z(2) = -0.4$ Bohr; (*f*) $z(2) = -0.4$ Bohr and $\varphi(2) = -0.02$ rad.

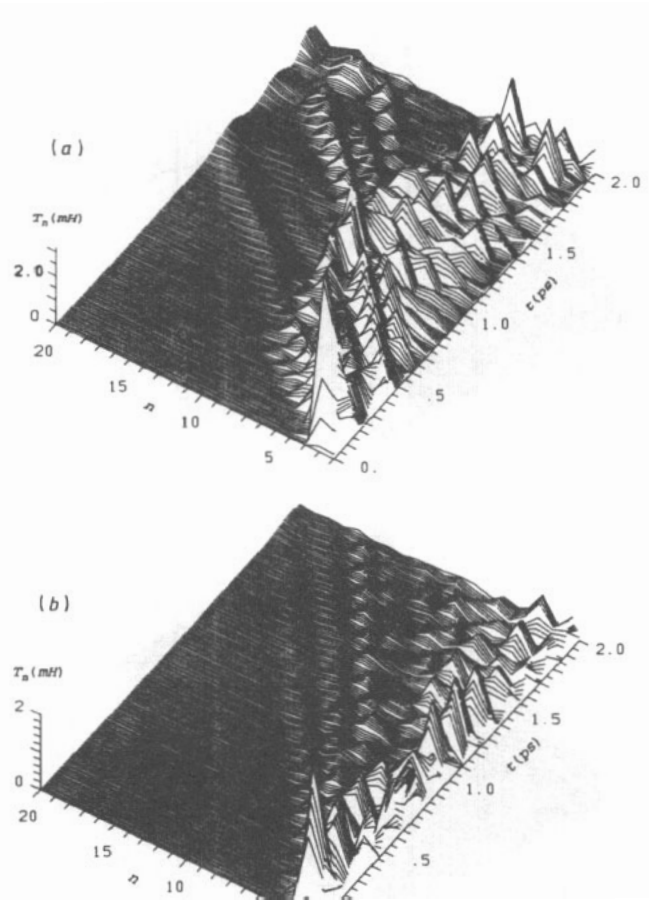


Figure 8. Time evolution of the local kinetic energy T_n as a function of site n and time for: (*a*) a double-stranded A–T stack with a reversed T–A pair in position 9; and (*b*) a double-stranded stack with alternating A–T and T–A base pairs; with an initial excitation of $z(2) = -0.4$ Bohr on both strands (open chain ends, angles frozen, geometry optimization prior to simulation, only one strand shown, second one identical to first one).

molecules in the backbone with the nucleotide bases was added. The sugar molecules were taken to be rigid during the simulations. In contrast to the previously described calculations now *none* of the degrees of freedom for the bases was frozen, because the additional C–N bonds between an N atom of the base and a C atom of the sugar stabilize the bases close to the measured crystal structure. In figure 9 we show the dynamics of a single-stranded adenine helix including the backbone potential. In this simulation the initial excitation was $z(10) = -0.1$ Bohr, the chain ends were kept fixed and the angles were allowed to change freely. Excitations in ϑ and in φ or combined ones show a similar behaviour. Thus we show only one example. Figure 9 shows that already after roughly 1.5 ps the initial excitation is dispersed through the whole chain and no solitary waves are formed.

Finally a double helix of hydrogen-bonded A–T base pairs including the backbone

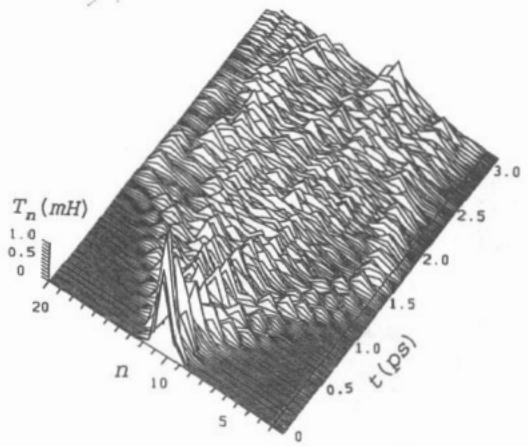


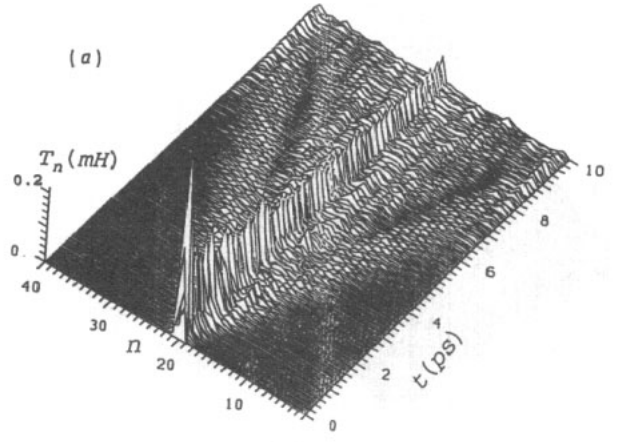
Figure 9. Time evolution of the local kinetic energy T_n as a function of time and site in a stack of adenine molecules stabilized by sugar as backbone (initial excitation $z(10) = -0.1$ Bohr, geometry optimization prior to simulation, fixed chain ends, all variables change freely).

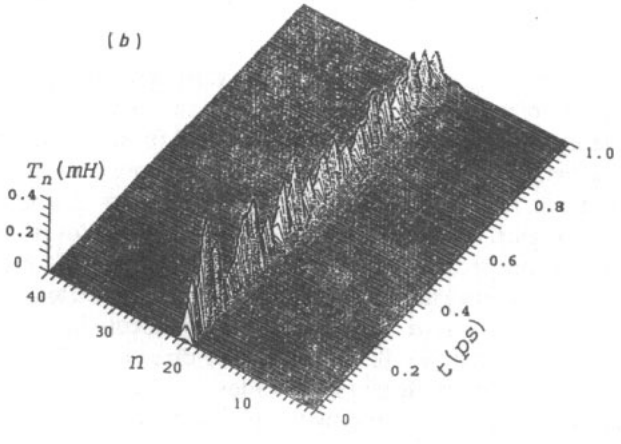
was treated. As a first step the geometry of the stack starting from the structure of B-DNA was optimized. The optimized geometry deviates from the experimental structure of B-DNA [29] for adenine by $\vartheta = 3.75°$, $\varphi = 0.13°$ and $z = -0.05$ Å and for thymine by $\vartheta = 2.41°$, $\varphi = 1.64°$ and $z = 0.11$ Å. These rather small deviations indicate that our model of B-DNA is rather realistic in the present stage, at least for static properties such as the equilibrium structure. Figure 10 shows the time evolution of the local kinetic energy in such a double helix starting from different initial excitations. Here all geometrical degrees of freedom are now free (none of them is frozen) and we used fixed chain ends.

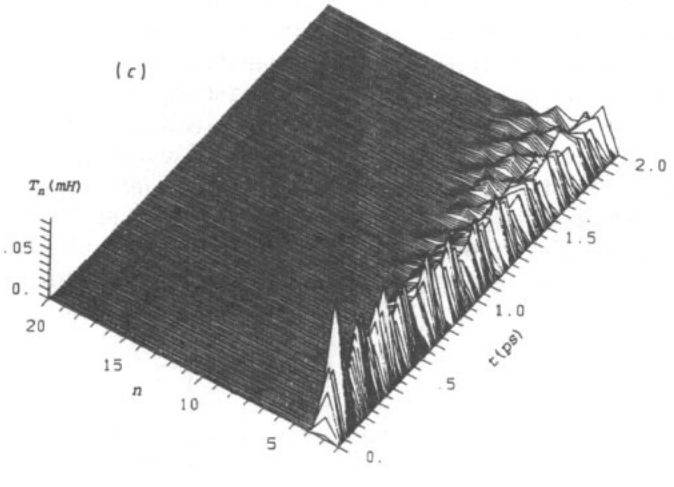
Figure 10(*a*) shows a case where only the adenine molecule in an A–T pair in the centre of the chain was displaced by -0.1 Bohr in the z direction. The dispersion of the excitation energy is slower than in the single-stranded case, which should be due to the coupling of the bases to the second strand by the two hydrogen bonds. For that reason the energy remains localized at the initial point for a longer time. To discuss the results in more detail the first picosecond of this simulation is enlarged in figure 10(*b*). It is obvious that a high-frequency vibration is excited that is not present in figure 9. This vibration is due to the excitation of a vibration of the C–N bond to the backbone together with the perturbation of the hydrogen bonds between A and T. It seems that the presence of an additional coupling in the system, which is much stronger than the one responsible for the transport, prevents the formation of solitary waves. For that reason a more flexible backbone (lowering the force between base and sugar) may support a solitary wave. In figures 10(*c*) and (*d*) we show the time evolution of initial displacements in the z direction of a whole hydrogen-bonded A–T base pair at site 2 for two excitation energies. We see that also in this case mainly a high-frequency vibration of the C–N bond is excited and the energy remains at the excitation site for a rather long time.

4. Conclusion

The basic aim of the present work was first of all to develop a realistic description of the dynamics of B-DNA and the investigation of the time evolution of conformational


(a)
T_n (mH)
0.2
0.
40
30
n
20
10
0
2
4
6
8
10
t (ps)


(b)
T_n (mH)
0.4
0.2
0
40
30
n
20
10
0
0.2
0.4
0.6
0.8
1.0
t (ps)


(c)
T_n (mH)
.05
0.
20
15
n
10
5
0.
.5
1.0
1.5
2.0
t (ps)

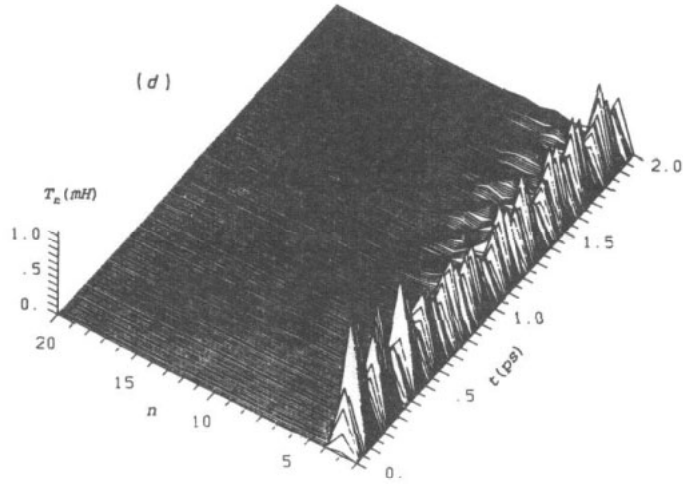


Figure 10. Time evolution of the local kinetic energy T_n as a function of time and site in a double helix of hydrogen-bonded A–T base pairs stabilized by sugar as backbone for the following initial excitations (all other conditions as in figure 9, only one strand shown): (*a*) $z(20) = -0.1$ Bohr (only the A molecule of the pair excited); (*b*) first picosecond of (*a*); (*c*) $z(2) = -0.05$ Bohr (A–T pair excited); (*d*) $z(2) = -0.2$ Bohr (A–T pair excited).

excitations of the nucleotide bases. With the help of our program we performed computer experiments and observed the formation of solitary waves under special conditions within our simplified model of DNA. This model could easily be extended to the description of, for example, melting phenomena in DNA, which up to now have been investigated with more approximate theories (e.g. two chains of harmonically coupled point masses, with Morse-type potentials between them to simulate the hydrogen bonds between bases) [30, 31]. For this purpose one can simply couple the sugar molecules, which are rigid in our model, by a harmonic force or again with the force-field model similar to the coupling between sugar and bases. In addition one would need one further degree of freedom, which describes displacements of the bases perpendicular to the helix axes. Work along this line is in progress in our laboratory.

Up to now within our theoretical model in all stacked (formamide, thioformamide, adenine, thymine and A–T) systems we have found conformational solitons. It seems that this is a widespread phenomenon and should exist, for example, in 7,7′, 8,8′-tetracyanoquinodimethane tetrathiofulvalene (TCNQ-TTF) stacks and similar systems, where solitons may even serve as charge carriers [32]. If an additional strong coupling besides the coupling necessary for transport (like the backbone in DNA) is present, the soliton is destroyed, as in our classical model with rigid sugar. However, we have worked out and applied a rather realistic model for the description of the dynamics of DNA. With the present version of the model we could not decide clearly on the presence or absence of conformational solitary waves in the base stacks of DNA; however, the model (and also the corresponding program) can be extended in a straightforward but tedious way as described below.

To be able to give a decisive answer whether solitons play a role in carcinogenesis or not, the model has to be extended. For example, a quantized oscillator (as the C–N bond

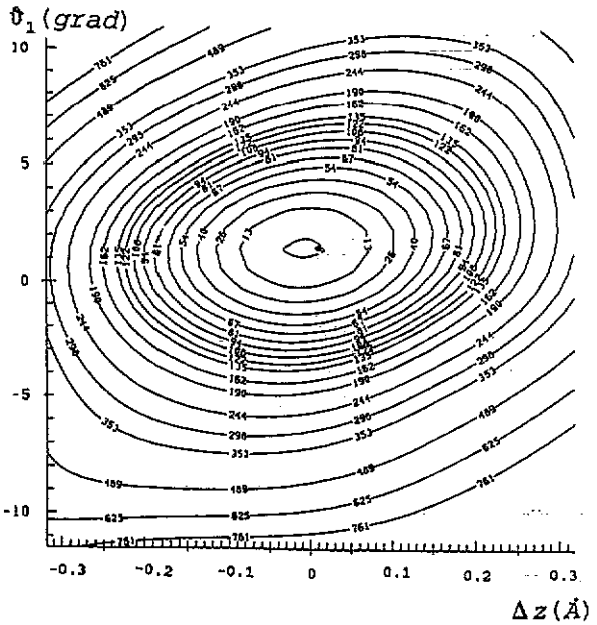


Figure 11. The $(\Delta z, \vartheta_1)$ potential energy surface for displacement of the adenine molecule of a single A–T pair in the centre of an adenine–thymine double helix with backbone (energy in meV).

stretching vibration) can be excited only with definite quanta of energy from one stationary state to another. In figure 11 we show the excitation energy of an adenine molecule in an A–T double helix as a function of the displacements Δz and ϑ_1. The figure shows that only displacements larger than 0.1 Å in Δz and about 5° in ϑ_1 cause an excitation energy larger than 150 meV, which is the vibrational (bond stretching) quantum of energy for a C–N bond known from infrared spectra. Up to this energy no stationary state of this oscillator exists. Thus in a first attempt to search for solitary waves in the base stacks of DNA, one can include quantum effects at least for the C–N bond stretching vibration in the equations of motion. However, one has to keep in mind that we are dealing with non-stationary states, which can take up any amount of energy.

Further one might search for other definitions of the degrees of freedom than those used in this work. For example, one can use a rotation of the bases around the C–N bonds instead of the coordinate ϑ. A rotation of this kind would also lead to a tilting of the bases as required, but would not excite the C–N stretch. In addition, one can think of a more flexible backbone to reduce the sugar–base coupling strength. A fit of a larger region of the potential surfaces would probably also be of importance. However, in this case one would have to expand the displacements in z into a $1/z^n$ series instead of a Taylor series and the angles into a Fourier series. However, these extensions of the model are beyond the scope of the present work, which is aimed more at the development of a basic model, which afterwards can be extended along the lines described above.

If the soliton mechanism were not possible in DNA, other long-range mechanisms have to be examined, like the rearrangement of the tertiary structure, the change of dispersion and polarization forces between DNA and protein chains due to charge transfer, etc [33].

Acknowledgments

The financial support of the 'Deutsche Forschungsgemeinschaft' (Project no. Ot/56-2) and of the 'Fonds der chemischen Industrie' is gratefully acknowledged.

References

[1] Kruskal M D and Zabusky N J 1963 *Princeton Plasma Phys. Lab. Annu. Rep.* MATT-Q-21, p 301
[2] Mikeska H J 1978 *J. Phys. C: Solid State Phys.* **11** L29
[3] Maki K 1980 *J. Low Temp. Phys.* **41** 327
[4] Aubry S 1976 *J. Chem. Phys.* **64** 3392
[5] Su W P, Schrieffer J R and Heeger A J 1979 *Phys. Rev. Lett.* **42** 1698
[6] Krumhansl J A and Alexander D M 1983 *Structure and Dynamics: Nucleic Acids and Proteins* ed E Clementi and R H Sarma (New York: Adenine Press) p 61
[7] Forinash K, Bishop A R and Lomdahl P S 1991 *Phys. Rev.* B **43** 10743
[8] Russel J S 1844 *Proc. R. Soc. Edinb.* p 319
[9] Mollenauer L F, Stolen R H and Gordon J P 1988 *Phys. Rev. Lett.* **45** 1095
[10] Nakajima K, Mizusawa H, Sawada Y, Akoh H and Takada S 1990 *Phys. Rev. Lett.* **65** 1667
[11] Su W P, Schrieffer J R and Heeger A J 1980 *Phys. Rev. Lett.* **42** 1698
[12] Ladik J 1990 *Physiol. Chem. Phys. Med. NMR* **22** 229
[13] Ladik J and Čížek J 1984 *Int. J. Quantum Chem.* **26** 955
[14] Hofmann D, Förner W and Ladik J 1988 *Phys. Rev.* A **37** 4429
[15] Hofmann D, Förner W, Otto P and Ladik J 1990 *J. Phys.: Condens. Matter* **2** 4081
[16] Förner W, Wang C L, Martino F and Ladik J 1988 *Phys. Rev.* B **37** 4567
[17] Förner W 1989 *Phys. Rev.* A **40** 6438
[18] Khang Y K and Jhon M 1982 *Theor. Chim. Acta* **61** 41
[19] Weiner S J, Kollman P A, Case D A, Singh U C, Ghio C, Alagona G, Profeta S and Weiner P 1984 *J. Am. Chem. Soc.* **106** 765
[20] Clementi E and Corongiu G 1982 *Int. J. Quantum Chem.* **QBS9** 213
[21] Otto P and Ladik J 1975 *Chem. Phys.* **8** 192
[22] Otto P and Ladik J 1977 *Chem. Phys.* **19** 209
[23] Otto P 1979 *Chem. Phys. Lett.* **62** (3) 538
[24] Otto P 1989 *J. Mol. Struct.* **188** 277
[25] Förner W, Otto P and Ladik J 1984 *Chem. Phys.* **86** 49
[26] Aida M and Nagata C 1982 *Chem. Phys. Lett.* **86** 44
[27] Gianolio L, Pavani R and Clementi E 1978 *Gazz. Chim. Acta* **108** 181
[28] Otto P PPTMCF 89 (unpublished)
[29] Arnott S, Dover S D and Wonnacott A J 1969 *Acta Crystallogr.* B **25** 2192
[30] Peyrard M and Bishop A R 1989 *Phys. Rev. Lett.* **62** 2755
[31] Forinash K, Bishop A R and Lomdahl P S 1991 *Phys. Rev.* B **43** 10743
[32] Förner W, Ladik J, Otto P and Martino F 1989 *Phys. Rev.* A **40** 6457
[33] Ladik J, Suhai S and Seel M 1978 *Int. J. Quantum Chem.* **QBS5** 35