

E. A. Kümmerle · E. Pomplun

A computer-generated supercoiled model of the pUC19 plasmid

Received: 3 May 2004 / Revised: 18 June 2004 / Accepted: 22 June 2004 / Published online: 6 August 2004
© EBSA 2004

Abstract DNA models have become a powerful tool in the simulation of radiation-induced molecular damage. Here, a computer code was developed which calculates the coordinates of individual atoms in supercoiled plasmid DNA. In this prototype study, the known base-pair sequence of the pUC19 plasmid has been utilized. The model was built in a three-step process. Firstly, a Monte Carlo simulation was performed to shape a segment chain skeleton. Checks on elastic energy, distance and unknotting were applied. The temperature was considered in two different ways: (1) it was kept constant at 293 K and (2) it was gradually reduced from 350 K to less than 10 K. Secondly, a special smoothing procedure was introduced here to remove the edges from the segment chain without changing the total curve length while avoiding the production of overshooting arcs. Finally, the base pair sequence was placed along the smoothed segment chain and the positions of all the atoms were calculated. As a first result, a few examples of the supercoiled plasmid models will be presented, demonstrating the strong influence of appropriate control of the system temperature.

Keywords DNA modelling · Monte Carlo simulation · Polyline smoothing · Supercoiled plasmid · Track structure calculations

Introduction

Track structure calculations have become a powerful tool in computational microdosimetric research during the past two decades. In the meantime, various computer codes now exist for simulating the transport of

different radiation particles with a broad kinetic energy range in liquid water and water vapour [see e.g. Terrissol and Vrigneaud (2001); Tomita et al. (1998); Paretzke (1987); for an overview see Nikjoo and Uehara (1994)]. These codes trace the history of individual particles from their starting points until the kinetic particle energy has been completely absorbed or at least below a threshold (< 10 eV). For each interaction with the irradiated matter's molecules (ionization, excitation etc.), the spatial coordinates, the energy transfer, secondary particles, etc., are recorded so that a track structure can be followed on the molecular level, displaying the characteristics of different types of radiation.

To utilize these tracks for the simulation of damage induction on the DNA molecule, in parallel to the code development, target models were generated. A parallel “doughnut” model of nucleosomal DNA was introduced by Charlton (1986) to investigate the range of low-energy Auger electrons released from incorporated ^{125}I . Two “doughnuts” in the form of curved solid cylinders of 2 nm in diameter represent the DNA helix turning around the histones. Furthermore, each cylinder was sliced into 80 discs representing the base pairs. The first linear DNA model which allowed a differentiation between base and sugar-phosphate volumes was created by the same author (Charlton and Humm 1988). A solid straight cylinder with a diameter of 2.3 nm was divided into three compartments; an inner cylinder of 1.0 nm diameter along the central axis is occupied by the bases. The two sugar-phosphates of a base pair were represented by 0.34-nm wide arches, each rotated by 36° for every successive base. The next step in this series was the linear helix model DNAMOD by Pomplun (1991), providing the coordinates of the individual atoms of the DNA molecule and thereby enabling for the first time the atomic resolution of a target model. A modification of DNAMOD led to a nucleosome structure with 144 base pairs turning the helix 1.75 times around the histone volumes (Pomplun and Terrissol 1994; Terrissol and Pomplun 1994). Also the positions of the water molecules in the inner and outer water sheath were

E. A. Kümmerle (✉) · E. Pomplun
Geschäftsbereich Sicherheit und Strahlenschutz,
Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: e.kuemmerle@fz-juelich.de
Fax: +49-2461-612166

included. Both of these models have become the kernel for higher-order models like chromatin fibre loops (Friedland et al. 1999).

The purpose of this paper is to provide a supercoiled plasmid DNA target model which can be used within track structure codes. Such molecules have often been used in radiobiological experiments, e.g. to investigate the damaging effectiveness of incorporated Auger electron emitters like ^{125}I (e.g. Kassis et al. 1999, 2000) or ^{111}In (Sahu et al. 1995). Supercoiled plasmid models have already been used for strand-break induction calculations for γ -ray or electron irradiation (Tomita et al. 1998; Watanabe and Saito 2002). Similar to the plasmid modelling procedure used there, our work is also based on the Monte Carlo algorithm of Vologodskii et al. (1992), using a segment chain skeleton. Additionally, we developed a subsequent smoothing algorithm to finally obtain a model structure without any edges. Furthermore, the trial changes of the segment chain during the Monte Carlo procedure were optimized on the basis of success statistics.

Generation of the supercoiled DNA target model

The supercoiled DNA model was built in three steps. (1) A closed chain consisting of straight segments of equal length (i.e. an equilateral polygonal knot) was varied under torsional and bending forces using a Monte Carlo simulation based on the method described by Vologodskii et al. (1992). (2) The resulting segment chain was smoothed to a quasi-continuously bent curve. (3) This curve was overlaid by a given DNA base pair sequence and the positions of all the atoms were calculated.

The programs were written in C++ using the expression and meta-template library *tmtem* (Petzold 2001–2003) to achieve the highest performance in vector and matrix calculations. The calculations were executed on a Linux PC with 1.8 GHz.

Monte Carlo simulation

Method

Following Vologodskii et al. (1992), the DNA molecule was modelled by a closed chain consisting of N rigid segments with a length of approximately $l = 10$ nm each (1/10 of the Kuhn statistical length). The exact segment length resulted from the total chain length $L = Nl$ (in the case of pUC19, $N = 91$ and $l = 9.962$ nm was chosen to obtain $L = 906.5$ nm).

The elastic energy of the chain is $E = E_b + E_t$ with the bending energy:

$$E_b = k_B T \alpha \sum_{i=1}^N \Theta_i^2 \quad (1)$$

where Θ_i is the angular displacement between the directions of the segments i and $i+1$, k_B is the Boltzmann constant, $T = 293$ K is the temperature and $\alpha = 2.403$ is the bending constant. The torsional energy is:

$$E_t = (2\pi^2 C/L)(\Delta Lk - Wr)^2 \quad (2)$$

where $C = 3 \times 10^{-28}$ J m is the torsional rigidity. Wr is the writhe of the segment chain and:

$$\Delta Lk = Lk - Lk_0 = \sigma Lk_0 \quad (3)$$

is the linking number difference between the actual linking number LK of the closed DNA molecule and the linking number $Lk_0 = L/z_0$ for relaxed DNA with the helix height $z_0 = 3.375$ nm (Chandrasekaran and Arnott 1989). σ is the specific linking number difference, also called superhelix density, which was chosen to be $\sigma = -0.04$ (typical σ values for supercoiled DNA molecules are between -0.05 and -0.07 ; see e.g. Bauer 1978). The writhe Wr was calculated according to Klenin and Langowski (2000) (method 2a).

The starting formation of the segment chain was a plain, regular polygonal knot. At each Monte Carlo step, a trial change was applied to the segment chain, the elastic energy of the new formation, E_{i+1} , was calculated and compared to the elastic energy of the previous formation, E_i . If $E_{i+1} < E_i$ or:

$$\exp[(E_i - E_{i+1})/(k_B T_M)] > \rho \quad (4)$$

where T_M is the Monte Carlo temperature (i.e. the temperature which controls the Monte Carlo simulation process) and ρ is a random number between 0 and 1, the trial change was accepted, otherwise it was rejected. A distance check was performed to consider excluded-volume effects. If the distance between any two (non-adjacent) segments was less than the effective DNA diameter $d = 3.5$ nm, the trial change was rejected. Finally, the Alexander polynomial was calculated for the segment chain using the program by Harris and Harvey (1999) to ensure that the new formation was unknotted; otherwise it was rejected.

Two types of trial changes were applied: (1) a crankshaft rotation of an arbitrarily chosen sub-chain around the axis given by the sub-chain end points by an angle ϕ and (2) a reptation move, i.e. the exchange of two arbitrarily chosen sub-chains of m and $m+1$ segments after appropriate deformation [see Vologodskii et al. (1992) for a more detailed explanation].

The following modifications were applied to the procedure described by Vologodskii et al. (1992):

1. The crankshaft rotation angle ϕ was not chosen completely arbitrarily but on the basis of success statistics. The ϕ range $-\pi$ to π was divided into 100 intervals of equal length. For each interval i the relative frequency f_i^c for accepting crankshaft rotation trial changes was calculated and continuously updated during the Monte Carlo simulation using an

exponential forgetting algorithm. The a priori probability p_i^c for trying a crankshaft rotation with a ϕ value belonging to the interval i was chosen to be $p_i^c \propto f_i^c \sqrt{\phi_i}$ where the factor $\sqrt{\phi_i}$ (with $\phi_i = \max(|\phi|)$ in the interval i) makes allowance for the fact that larger ϕ values are more effective in changing the segment chain formation than small ones. At the beginning of the Monte Carlo simulation the p_i^c values were set to the same start value for all intervals. Within the selected interval, the applied ϕ value was chosen randomly.

2. The size m of the smaller reptation move sub-chain was not fixed to 3 but varied between 2 and 9. Here, too, success statistics were built. The a priori probability p_j^r for trying a reptation move with $m=j$ was chosen to be $p_j^r \propto f_j^r$ with the relative frequency f_j^r for accepting such reptation move trial changes. At the beginning, all p_j^r values were initialized with the same value.
3. The a priori probabilities for crankshaft rotations and reptation moves were chosen to be 2:1.

Progression of the Monte Carlo simulations

Two versions of the Monte Carlo procedure were performed: (1) using a constant Monte Carlo temperature, $T_M = 293$ K, during the entire simulation and (2) starting with $T_M = 350$ K and reducing T_M gradually to a final value of typically less than 10 K. Here, T_M (cf. Eq. 4) was reduced by the factor 0.9 whenever the elastic energy averaged over an interval of 1000 successful trial changes was greater than 0.9 times the averaged elastic energy in the preceding interval. The final chains thus represent (1) a snapshot of a DNA molecule subjected to thermal deformations and (2) a frozen-in DNA molecule (note: the temperature $T = 293$ K in Eq. 1 was never changed).

Figures 1 and 2 show examples of the evolution of the elastic energy for both versions with $N=91$. The number of Monte Carlo steps was 10^6 each. In the case of constant T_M , the elastic energy already reaches its final fluctuation range after approximately 5×10^4 Monte Carlo steps (Fig. 1). In the case of T_M being reduced, the decrease of the elastic energy was controlled by the decrementation of T_M (Fig. 2).

Figure 3 shows the success statistics for crankshaft rotation and reptation move trial changes for constant T_M . The acceptance probability for crankshaft rotations decreased strongly with increasing angle $|\phi|$. The most frequently accepted reptation moves were those with the smaller subchain sizes, $m=4$ or $m=5$.

Smoothing of the segment chain

The requirements for the smoothing algorithm were (1) to approximately conserve the total length of the curve and (2) not to produce overshooting arcs. The widely

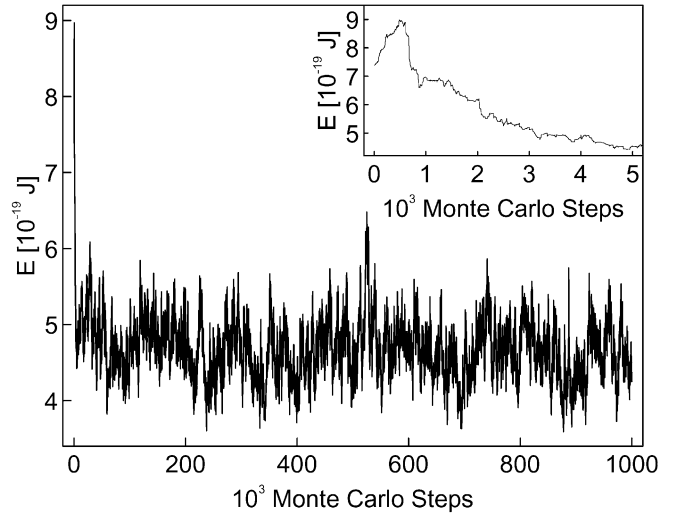


Fig. 1 Evolution of the elastic energy E during the Monte Carlo procedure for a segment chain with $N=91$ and $\sigma=-0.04$ with constant $T_M=293$ K

used spline functions, for example, do not fulfil any of these criteria, so the following procedure was applied.

Smoothing was done iteratively and, in every iteration cycle, each chain segment was replaced by two segments of approximately half the length. At every iteration, first the bisecting vectors \mathbf{a}_i of the angles between two adjacent segments \mathbf{s}_{i-1} and \mathbf{s}_i and the perpendicular bisecting planes B_i of the segments \mathbf{s}_i were calculated (see Fig. 4). Then the intersection points \mathbf{P}_i^1 and \mathbf{P}_i^2 of a plane B_i with the two neighbouring vectors \mathbf{a}_i and \mathbf{a}_{i+1} and the lines \mathbf{b}_i^1 and \mathbf{b}_i^2 in the plane B_i through the points \mathbf{P}_i^1 and \mathbf{P}_i^2 , respectively, and crossing the segment \mathbf{s}_i were determined.

In the next step, provisional new segments $\tilde{\mathbf{s}}_{2i}$ and $\tilde{\mathbf{s}}_{2i+1}$ were constructed by calculating the base side of the isosceles triangles with the equal sides \mathbf{a}_i and \mathbf{b}_i^1 and \mathbf{b}_i^2

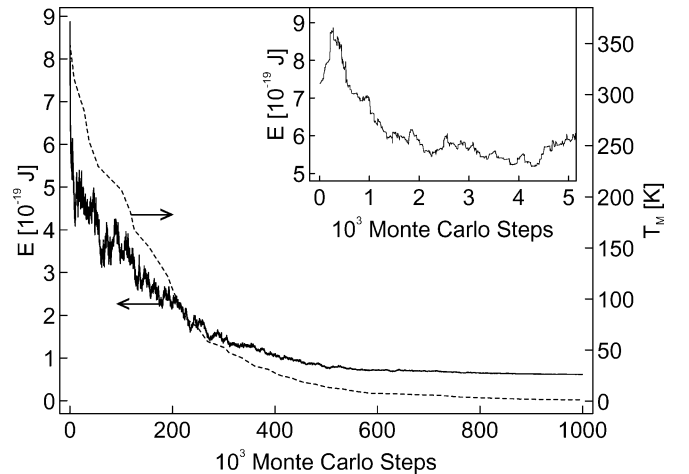


Fig. 2 Evolution of the elastic energy E during the Monte Carlo procedure for a segment chain with $N=91$ and $\sigma=-0.04$ where T_M (dashed line) was gradually reduced starting with $T_M=350$ K

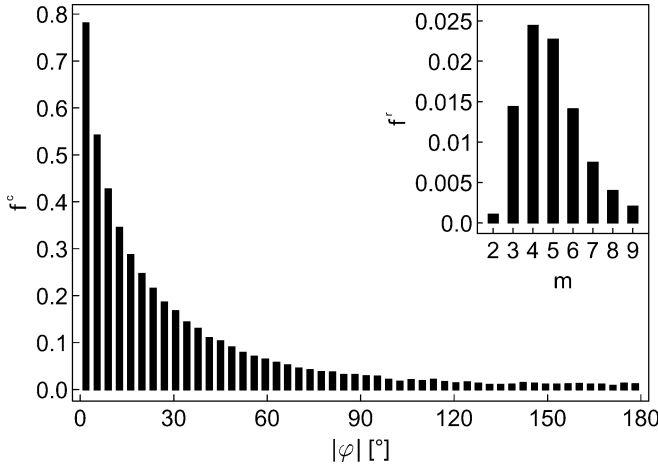


Fig. 3 Success statistics for a Monte Carlo procedure with $N=91$, $\sigma=-0.04$ and constant $T_M=293$ K. f_c and f_r are the relative frequencies for accepting crankshaft rotation (angle ϕ) and reptation move (size m of the smaller subchain) trial changes

and \mathbf{a}_{i+1} , respectively, and with the length of the base side equal to half the length of the old segment \mathbf{s}_i . Generally, the end points of adjacent provisional segments $\tilde{\mathbf{s}}_j$ do not coincide, i.e. the segments are not connected. The final new segments \mathbf{s}'_j were therefore determined by taking the centre points of each two end points of the $\tilde{\mathbf{s}}_j$ which have to be connected as the corners of the new segment chain.

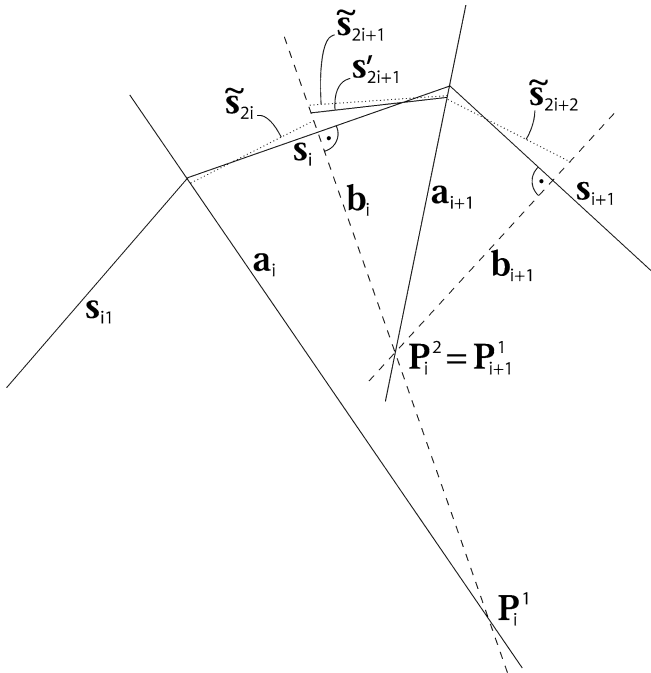


Fig. 4 The smoothing procedure illustrated in two dimensions. Contrary to the general three-dimensional case, the points \mathbf{P}_i^2 and \mathbf{P}_{i+1}^1 coincide. Furthermore, the lines \mathbf{b}_i^1 and \mathbf{b}_{i+1}^2 coincide in two dimensions, labelled \mathbf{b}_i in the figure. This line also represents the plane B_i

In the case of a plane regular polygon, all the provisional new segments $\tilde{\mathbf{s}}_j$ are connected, so this procedure exactly conserves the length of the curve. In the general case of an irregular polygonal knot, the total length of the curve is typically reduced by less than 1% and the length of individual segments may typically vary from $\sim 93\%$ to $\sim 102\%$ after five iteration steps. The small change of the total length was finally compensated by rescaling the length unit.

At every iteration step a distance check was performed and at the end of the smoothing procedure a final knot check was made. After five iteration steps the maximum angle between two adjacent segments was reduced from typically more than 90° to less than 10° . A final calculation of the writhe showed that it was changed by typically less than 0.5% due to the smoothing procedure. Figure 5 shows as an example a detail of a smoothed segment chain.

DNA sequence overlay

Now a given sequence of DNA base pairs, namely that of the pUC19 plasmid (Yanisch-Perron et al. 1985), was placed along the smoothed segment chain. While following that curve the coordinate system in which the positions of the DNA base atoms were defined was moved ahead, considering the twist $\Delta Tw = \Delta Lk - Wr$ of the curve. Its orientation was rotated continuously from each segment of the smoothed chain to the next one by linear interpolation. Figure 6 illustrates the result.

Results and conclusions

Figures 7 and 8 show examples of generated supercoiled DNA model structures for the plasmid pUC19 where the Monte Carlo temperature was reduced from $T_M=350$ K to 1.5 K or $T_M=293$ K was kept constant. In the first case the resulting structure is formed in a very orderly

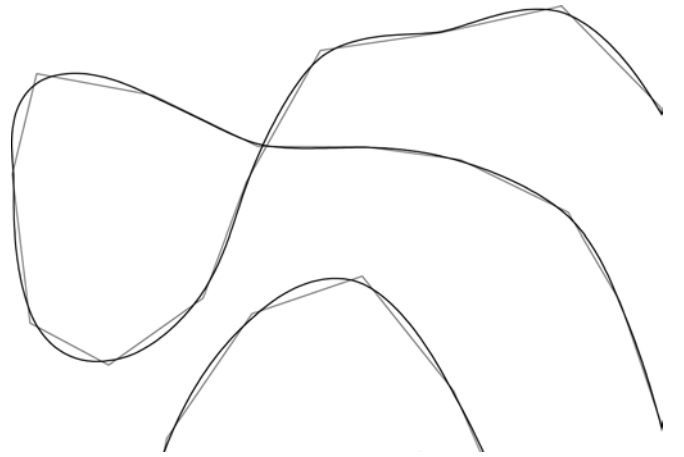
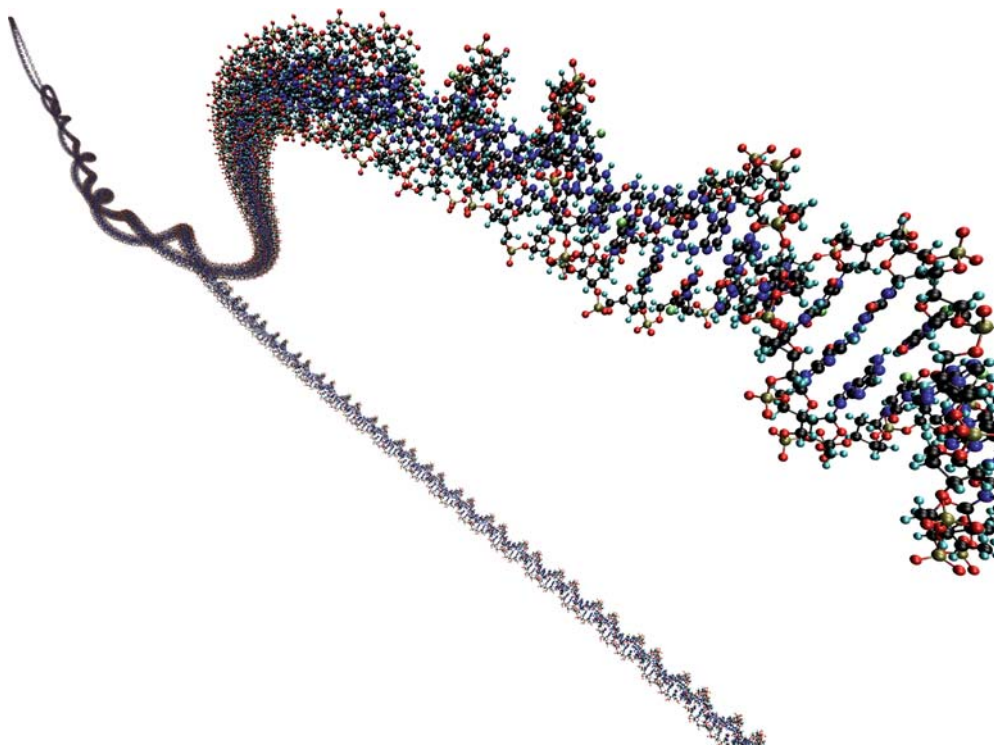


Fig. 5 Example of a detail of an initial segment chain (grey) and the associated smoothed segment chain (black)

Fig. 6 Example of a detail of a DNA model structure generated by overlaying a DNA sequence on a smoothed segment chain (the same model structure as in Fig. 7). All molecule visualizations are generated using the molecular visualization program VMD (Humphrey et al. 1996)



manner, while in the second case the structure is heavily distorted by thermal deformations, leading to an irregular shape. The writhe of the thermally deformed structure, $Wr = -8.2$, is somewhat smaller in its absolute value than that of the frozen-in structure, $Wr = -9.9$. This clearly demonstrates that the temperature is a key parameter in the simulation of realistic physiological plasmid forms.

A particular problem of using a DNA plasmid model for track structure calculations is the great variety of

possible geometric conformations at ambient temperature. The actual form of a plasmid molecule is strongly influenced by thermal motions. Track structure calculations can be applied only on a very limited number of sample conformations. To find a small set of conformations which can be recognized as representative will be the next step in providing models for DNA damaging investigations. [Nevertheless, it is justifiable to use static plasmid models in track structure calculations because the form of a plasmid molecule changes on a time scale

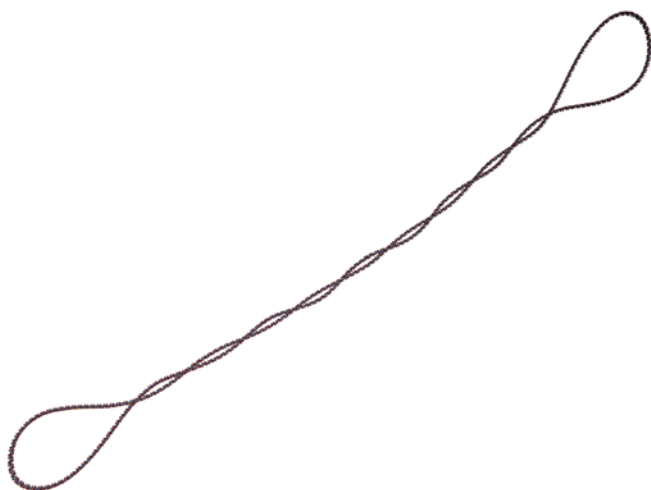


Fig. 7 Supercoiled DNA model structure for pUC19 (2686 base pairs) and the superhelix density $\sigma = -0.04$ generated by reducing the Monte Carlo temperature gradually from $T_M = 350$ K to 1.5 K. The writhe is $Wr = -9.9$

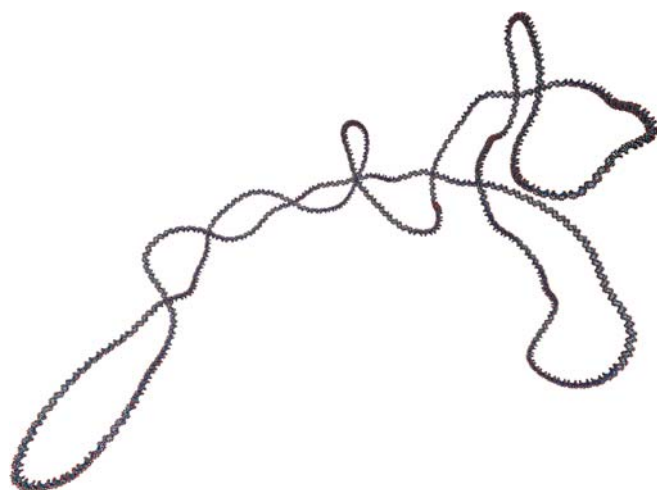


Fig. 8 Supercoiled DNA model structure for pUC19 (2686 base pairs) and the superhelix density $\sigma = -0.04$ generated with a constant Monte Carlo temperature $T_M = 293$ K. The writhe is $Wr = -8.2$

of about 10^{-6} s (Jian et al. 1998), which is very slow compared to the duration of the chemical phase of the radiation action chain of about 10^{-8} s.]

The successful modelling of the pUC19 plasmid shown here now provides the ability to also build models for other plasmids where the base pair sequence is known. As a further candidate, the *Escherichia coli* plasmid pBR322 with 4361 base pairs is under consideration, being a biomolecule which is also very frequently used in radiation biology.

References

- Bauer WR (1978) Structure and reactions of closed duplex DNA. *Annu Rev Biophys Bioeng* 7:287–313
- Chandrasekaran R, Arnott S (1989) The structures of DNA and RNA helices in oriented fibers. In: Landolt-Börnstein – group VII, biophysics 1B, crystallographic and structural data II. Springer, Berlin Heidelberg New York, pp 31–170
- Charlton DE (1986) The range of high LET effects from ^{125}I decays. *Radiat Res* 107:163–171
- Charlton DE, Humm JL (1988) A method of calculating initial DNA strand breakage following the decay of incorporated ^{125}I . *Int J Radiat Biol* 53:353–365
- Friedland W, Jacob P, Paretzke HG, Merzagora M, Ottolenghi A (1999) Simulation of DNA fragment distributions after irradiation with photons. *Radiat Environ Biophys* 38:39–47
- Harris BA, Harvey SC (1999) Program for analyzing knots represented by polygonal paths. *J Comput Chem* 20:813–818 (now that uracil.cmc.uab.edu has been shut down, the program is available at <http://rumour.biology.gatech.edu/Publications/1998–2000.shtml>)
- Humphrey W, Dalke A, Schulten K (1996) VMD – visual molecular dynamics. *J Mol Graphics* 14:33–38 (<http://www.ks.uiuc.edu/Research/vmd>)
- Jian H, Schlick T, Vologodskii A (1998) Internal motion of supercoiled DNA: Brownian dynamics simulations of site juxtaposition. *J Mol Biol* 284:287–296
- Kassis AI, Harapanhalli RS, Adelstein SJ (1999) Comparison of strand breaks in plasmid DNA after positional changes of Auger electron-emitting Iodine-125. *Radiat Res* 151:167–176
- Kassis AI, Walicka MA, Adelstein SJ (2000) Double-strand break yield following I-125 decay – effects of DNA conformation. *Acta Oncol* 39:721–726
- Klenin K, Langowski J (2000) Computation of writhe in modeling of supercoiled DNA. *Biopolymers* 54:307–317
- Nikjoo H, Uehara S (1994) Comparison of various Monte Carlo track structure codes for energetic electrons in gaseous and liquid water. In: Varma MN, Chatterjee A (eds) *Computational approaches in molecular radiation biology*. Plenum, New York, pp 167–185
- Paretzke HG (1987) Radiation track structure theory. In: Freeman GR (ed) *Kinematics of nonhomogeneous processes*. Wiley, New York, pp 89–170
- Petzold O (2001–2003) Tiny vector matrix library using expression templates. <http://tvmet.sourceforge.net>
- Pomplun E (1991) A new DNA target model for track structure calculations and its first applications to I-125 Auger electrons. *Int J Radiat Biol* 59:625–642
- Pomplun E, Terrissol M (1994) Low-energy electrons inside active DNA models: a tool to elucidate the radiation action mechanisms. *Radiat Environ Biophys* 33:279–292
- Sahu SK, Kassis AI, Makrigiorgos GM, Baranowska-Kortylewicz J, Adelstein SJ (1995) The effects of indium-111 decay on pBR322 DNA. *Radiat Res* 141:193–198
- Terrissol M, Pomplun E (1994) A nucleosome model for the simulation of DNA strand break experiments. In: Varma MN, Chatterjee A (eds) *Computational approaches in molecular radiation biology*. Plenum, New York, pp 243–250
- Terrissol M, Vigneaud JM (2001) Analogue Monte Carlo to model radiation induced DNA damage. In: Kling A, Barao F, Nagakawa M, Távora LMN, Vaz P (eds) *Advanced Monte Carlo for radiation physics, particle transport simulation and applications*. Springer, Berlin Heidelberg New York, pp 261–266
- Tomita H, Kai M, Kusama T, Ito A (1998) Monte Carlo simulation of DNA strand-break induction in supercoiled plasmid pBR322 DNA from indirect effects. *Radiat Environ Biophys* 36:235–241
- Vologodskii AV, Levene SD, Klenin KV, Frank-Kamenetskii M, Cozzarelli NR (1992) Conformational and thermodynamic properties of supercoiled DNA. *J Mol Biol* 227:1224–1243
- Watanabe R, Saito K (2002) Monte Carlo simulation of strand-break induction on plasmid DNA in aqueous solution by monoenergetic electrons. *Radiat Environ Biophys* 41:207–215
- Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33:103–119