Filip Lankaš
German Cancer Research Centre (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany, and J. Heyrovsky Institute and Centre for Complex Molecular Systems and Biomolecules, Dolejškova 3, 182 23 Praha 8, Czech Republic

Received 15 August, 2003
accepted 15 August 2003

DNA Sequence-Dependent Deformability—Insights from Computer Simulations

Abstract: *The article reviews some recent developments in studying DNA sequence-dependent deformability, with emphasis on computer modeling. After a brief outline of available experimental techniques, we proceed to computational methods and focus on atomic-resolution molecular dynamics (MD) simulations. A sequence-dependent local (base-pair step) force field inferred from MD is compared with force fields obtained by other techniques. Various methods for establishing global (flexible-rod) DNA elastic constants are reviewed, including an approach based on atomic resolution MD. The problem of defining the global deformation variables, as well as the question of anisotropy and nonlocal effects, are discussed. As an example, both local and global deformability calculations from atomic-resolution MD of* Eco*RI dodecamer are presented.* © 2003 Wiley Periodicals, Inc. Biopolymers 73: 327–339, 2004

Keywords: *DNA flexibility; elastic properties; persistence length;* Eco*RI dodecamer; AMBER*

INTRODUCTION

In recent years we have witnessed an increasing number of experimental and theoretical studies of DNA structure, dynamics, and other physicochemical properties. This reflects the central biological role of DNA as a molecule carrying genetic information. DNA is much more than a chain of letters: a given sequence of nucleotides, together with a specific environment (solution conditions, mechanical stress, bound proteins), determine a unique three-dimensional structure, dynamic behavior, and mechanical properties of DNA. Evidence is accumulating that preferential binding sequences for proteins are determined not only by the possibility of specific chemical contacts between a protein and DNA, but also by suitable geometrical arrangement of the DNA fragment (e.g., curvature) and/or by its propensity to adopt a deformed confor-

Correspondence to: Filip Lankaš; email: filip.lankas@jh-inst.cas.cz

Present address: Bernoulli Institute for Mathematics EPFL (Swiss Federal Institute of Technology) CH-1015 Lausanne Switzerland

Contract grant sponsor: Ministry of Education of the Czech Republic (MECR), Volkswagen Foundation, and NATO. Contract grant number: LN00A032 (MECR) and LST.CLG.977846 (NATO)

Biopolymers, Vol. 73, 327–339 (2004)

mation facilitating the protein binding. As an example, the preference of some DNA sequences for wrapping around histones to form nucleosomes is related to DNA curvature and deformability. Nucleosomes constitute basic units of higher order DNA organization in the eukaryotic cell nucleus where distant regions of the genome may appear in close contact and influence each other via specific DNA spatial arrangement. Recent applications of DNA in the field of nanotechnology also rely, among other features, on sequence-dependent DNA structure and mechanical properties.

The aim of this article is to review some of the recent developments in studying DNA sequence-dependent deformability, with emphasis on computer modeling. We will concentrate on methods rather than applications. After a brief outline of several available experimental techniques, we will proceed to computational methods and focus on those based on atomic-resolution molecular dynamics. Since DNA exhibits a hierarchy of structures, its properties can be studied on various levels of detail (length scales). A problem then arises how to proceed consistently from a lower length scale to a higher one, i.e., how to devise a suitable coarse-graining procedure. We will discuss several approaches aimed at connecting the "local" (base-pair step) and "global" (flexible-rod) levels. A model system, the *Eco*RI dodecamer, will be presented as an illustrative example. Finally, we will try to outline some conclusions and perspectives.

We also refer the reader to the review by Olson and Zhurkin,[1] which covers similar topics.

EXPERIMENTAL APPROACHES

A range of experimental methods has been used to determine DNA mechanical properties. Each of them is able to deal with particular aspects of DNA deformability at a certain time scale.

Various micromanipulation techniques use long DNA fragments (tens of micrometers). The pioneering studies of Smith et al.,[2,3] Bustamante et al.,[4] Cluzel et al.,[5] and Strick et al.[6] provided for the first time a more complete picture of DNA mechanical properties: DNA is extensible (with macroscopic stretch modulus around 1100 pN) and exhibits twist–stretch coupling. The study of Baumann et al.[7] of ionic effects on the DNA elasticity beautifully shows that, contrary to a homogeneous elastic rod,[8] the stretch modulus and bending rigidity of DNA are not proportional: while bending persistence length decreases with increasing ionic strength, stretch modulus increases.

The method of cyclization, proposed already long time ago,[9,10] has been recently developed further by Crothers and co-workers and used by them to study sequence-dependent DNA structure and deformability. The method relies on the determination of the J factor, defined as the ratio of equilibrium constants for ligatable monomers and dimers of the fragment under study. It is equivalent to the concentration of one end of the fragment in the vicinity of the other end. The inclusion of a curved region in the construct (by means of phased A-tracts) enables one to deconvolute permanent bends from dynamic bending fluctuations. Roychoudhury et al.[11] used the method to study global structure and mechanical properties of the TATAAACGCC sequence, a strong nucleosome positioning motif.[12] Later, Nathan and Crothers[13] studied changes in mechanical properties of *Eco*RI DNA upon methylation. Part of the method also included Monte Carlo simulations and multivariable parameter fitting that made the data analysis rather demanding. However, Zhang and Crothers[14] have recently proposed a semianalytical treatment of the problem based on an iterative search for the minimum energy configuration of circular DNA and subsequent evaluation of thermodynamic quantities under harmonic approximation. This approach is mainly applicable to short DNA fragments where thermal fluctuations are relatively small. The model is efficient and capable of dealing with a base-pair-level inhomogeneity in bending and flexibility, thus paving the way for high-throughput studies of sequence-dependent deformability. Unfortunately, its sensitivity to changes in torsional modulus and bending anisotropy seems to be rather limited.

A very recent report by Zhang and Crothers[15] announces a comprehensive high-throughput method for detection of DNA bending and flexibility based on cyclization. It combines their statistical mechanical treatment mentioned above with two other improvements: a combinatorial method for making DNA constructs and a fluorescence method for monitoring the kinetics of ligation.

Another class of methods is based on observing the overall dynamics of a DNA fragment using a probe attached to it. Fluorescence polarization anisotropy (FPA)[16–18] analyses a fluorescence signal from an intercalated dye. It has been recently used by Pedone et al.[19] to measure torsional constant of ten 27-bp (base pair) fragments of different sequence. Schurr and co-workers[20] studied dynamic bending rigidity of a 200-bp DNA by transient polarization grating (TPG). Whereas time-resolved FPA spans the time scale of DNA movements up to ca. 120 ns, the TPG technique covers the range from 20 ns to 10 μs.

Schurr and co-workers established the bending persistence length of ca. 200 nm. Given the estimate of 137 nm for the static persistence length and 50 nm for the total one, they conclude that apart from the static bends and rapidly relaxing bending fluctuations, slowly relaxing bending transitions (outside the time range of their method) must exist as well. In a series of studies, Robinson and co-workers applied a site-specific electron paramagnetic resonance (EPR) spin probe technique to study DNA dynamics and deformability. Studying fragments 14–100 bp long,[21] they measured a dynamic bending persistence length (on a submicrosecond time scale) of 150–170 nm, stressing again the possible presence of slowly fluctuating bends. Later, they found that the alternating polyAT inserts are 20% more flexible than the control sequence,[22] suggesting exceptional bending flexibility of the AT repeats. They recently applied the method to a systematic study of sequence-dependent isotropic bending rigidity on the dinucleotide level[23] by investigating forty 50-bp fragments of different sequences and fitting the data to dinucleotide models. Such an approach results in a mesoscopic force field for dinucleotides, aimed at predicting mechanical properties of a sequence of any length. However, they found that several sequences in their pool cannot be reasonably fitted to a dinucleotide model.

Another dinucleotide force field has been proposed by De Santis and co-workers,[24] recently in a slightly modified form.[25] They assume that the bending and twisting rigidity of a generic DNA is modulated on the base-pair step level according to the dinucleotide melting temperature. They argue that the mean amplitude of bending and twisting fluctuations at which DNA falls apart (melts) is approximately the same for all steps, and that the harmonic approximation remains valid up to temperatures close to the melting one: since the mean fluctuation amplitude is then proportional to T/C, where T is temperature and C a force constant, higher melting temperature indicates proportionally higher stiffness. Their scale is based on the melting data by Gotoh and Tagashira.[26]

A lot of effort has been put into the analysis of databases of DNA x-ray structures. In most of the studies, the term "flexibility" stands for the root mean square deviation of a particular conformational parameter among different instances of a given motif (e.g., a dinucleotide) in the database. The conformations of a dinucleotide differ due to external perturbations such as crystal packing forces, sequence context, or binding to proteins in protein–DNA complexes. If many such perturbations (each with a certain probability distribution) act independently, then the resulting distortion will have a Gaussian distribution—a consequence of the central limit theorem of statistics. Using the known relationship between the covariance matrix of a multidimensional Gaussian distribution and the stiffness matrix,[27] one can in principle determine the harmonic force constants including all coupling terms. Olson and co-workers[28] used this approach to construct a harmonic dinucleotide-level force field with respect to the six base-pair step parameters (twist, tilt, roll, shift, slide, rise) with all coupling terms (twist–roll, etc.). This force field is more detailed than any of those mentioned above. However, they all share the property of being relative, typically with one free parameter that has to be calibrated to get actual values of the force constants. It may be the "generic" persistence length, or as for the force field of Olson and co-workers, the effective temperature of the statistical ensemble. It should be stressed that the effective temperature is a purely statistical property and is not related to the actual, physical temperature of the crystals.

NMR can provide information about how DNA sequence modulates various aspects of dynamic behavior,[29,30] e.g., base-pair lifetimes and dissociation constants for opening.[31]

COMPUTATIONAL METHODS

Most computational approaches used to study DNA deformability are based either on potential energy calculations or on atomic-resolution molecular dynamics.

In a pioneering study, Zhurkin et al.[32] used a specially developed force field to investigate DNA anisotropic flexibility by calculation of potential energy changes upon deformation. The energy minimization program JUMNA, developed by Lavery and co-workers, is based on internal variable representation of DNA and uses implicit solvent modeled by a sigmoidal dielectric function. It has been widely used to study mechanics and energetics of large conformational transition in DNA. More detailed discussion of these results is outside the scope of the present study—see the review of Lafontaine and Lavery.[33]

In a series of studies on modeling DNA sequence-dependent structure and flexibility, Packer and Hunter first used a dinucleotide force field with backbone treated as a semiflexible rod.[34] The only free parameters are shift and slide; all the others are optimized. The calculated energy minima are in partial agreement with x-ray structures from a database assembled by the authors. The agreement is improved upon passing to tetranucleotides[35] where a new term is added to the energy expression. This new term is a phenome-

nological penalty function that should reflect the correlation of slide and shift in neighboring steps observed in the database. The harmonic deformability with respect to slide of the central step is calculated by fitting a quadratic function to grid points around the minimum. Recently, the authors moved on to study 8–12 bp oligomers from their database.[36] Here, yet another penalty function is added, reflecting the grouping of slide in the database around a certain value. Thus, their approach combines energy function derivation based on a physical concept with phenomenological data fitting.

Contrary to potential energy calculations, atomic-resolution molecular dynamics (MD) simulations treat the system at nonzero temperature, thus enabling it to cross small potential barriers and exhibit thermal fluctuations that can serve as a valuable source of information. MD dynamics simulations have come of age and are widely used to study both proteins and nucleic acids—see recent reviews.[37–39]

How to use molecular dynamics for assessing free energy changes upon deformation? Various approaches are based on a formula for the probability distribution of fluctuations, first used by Einstein around 1910.[27] Let x be a fluctuating variable related to a body (e.g., a molecule); we assume that the mean value $\bar{x}$ has already been subtracted from x, so that $x=0$ at equilibrium. Let S be the entropy of the whole system (body plus thermal reservoir) formally considered as a function of the exact values of x. Then the probability distribution of x is given by $w(x)=\text{const.}\exp[S(x)]$. Instead of the full entropy S, one can use its deviations from the equilibrium value. This quantity can be proved to coincide with the minimum (i.e., equilibrium) work necessary to change x from equilibrium to a given value, divided by kT. At constant temperature and pressure, this work is equal to the change ΔG of the free energy of the body. We thus have

$$w(x) = \text{const.}\exp(-\Delta G(x)/kT) \qquad (1)$$

or $\Delta G(x)=-kT\ln w(x)+\text{const}$. At equilibrium, $x = 0$ and $\Delta G(x)=0$, which gives the value of the constant in the last equation, yielding

$$\Delta G(x) = -kT \ln(w(x)/w(0)) \qquad (2)$$

These formulas are valid for any fluctuations, big or small, and require no assumptions about the functional form of ΔG. They hold also for the case of multiple variables. The probability distribution $w(x)$ can be easily inferred from MD simulations and formulas (1) or (2), then applied to obtain the free energy changes.

One way to do so consists in restraining the variable x by an additional harmonic potential ("umbrella potential"). Short MD runs are performed for different values of x, yielding a biased probability distribution. The biased distribution is then corrected for the umbrella potential to yield the unbiased distribution $w(x)$, from which the free energy values are extracted, using a version of the weighted histogram method.[40,41] This approach has recently been applied by several groups for establishing free energy changes associated with DNA base-pair opening[42–44] or conformational transitions in DNA backbone.[45]

Another possibility is to run an unrestrained MD and use Eq. (2) directly to get free energy values at discrete points (corresponding to chosen binning intervals for constructing $w(x)$ from MD data), which can then be fitted to a smooth function. Due to limited MD sampling, the results are restricted to the vicinity of a particular minimum, but the fitting function can be anharmonic. Osman and co-workers recently used this approach to investigate the energetics of base-flipping in damaged DNA.[46] They sampled the vicinity of two different states connected by a reaction coordinate (linear combination of bending and base opening) and fitted the data on the coordinate by sixth-order polynomial expansion. The force constants were calculated from the second derivatives of the fitted function in the minima. Since the data for both states were included, the free energy barrier between them could be estimated as well (some caution is necessary here, since the MD does not sample a vast transitional region between the two states). The free energy of one state with respect to the other had to be defined by means of a suitable thermodynamic cycle, because the simulations of the two states do not share any common configuration. In another study,[47] Osman and co-workers constructed a two-dimensional free-energy surface of DNA helicoidal parameters (inclination and tip). Obviously, the fitting would become increasingly difficult as the number of dimensions (degrees of freedom) would increase.

This difficulty can be overcome if one assumes that the free energy function is quadratic (harmonic). If x_i, $i = 1 \ldots N$, are the conformational variables (with mean values already subtracted), the most general quadratic free energy expression is $\Delta G=\frac{1}{2}\Sigma_{i,j=1}^{N} F_{ij}x_i x_j$, where F stands for the stiffness matrix (matrix of force constants). Equation (1) can then be solved analytically,[27] yielding that the inverse of the stiffness matrix is proportional to the correlation matrix of the conformational variables:

$$\langle x_i x_j \rangle = kT(F^{-1})_{ij} \quad (3)$$

The correlation matrix can be obtained from unrestrained MD and then inverted to get the stiffness matrix F. If there are no couplings between conformational variables, ΔG is simply a sum of squares and formula (3) reduces to the well-known "equipartition principle." The latter has been used in an early study by Bruant et al.[48] to calculate bending, twisting, and stretching stiffness of two 15-bp oligomers containing strong and weak TATA binding sites.

Lankas et al.[49] used eq. (3) to establish sequence-dependent global (flexible-rod) elasticity of selected DNA oligomers. They used four conformational variables: total fragment length, total twist, and two bending angles (into the grooves and into the backbone of the fragment center). Later, they applied the method as part of a study on the effect of exocyclic groups on DNA properties.[50] Recently, they used the same approach to construct a sequence-dependent force field with respect to the six base-pair step conformational parameters (roll, tilt, twist, shift, slide, rise).[51]

A theoretical framework for extracting parameters for base-pair level models of DNA from MD has been developed by Gonzalez and Maddocks.[52] Instead of Eq. (1), which is based on macroscopic considerations, they adopted a microscopic approach in which the standard canonical measure is proportional to the exponential of the hamiltonian. The change of variables from canonical to noncanonical ones (i.e., rotational parameters such as twist) leads to a jacobian J in the expression of the measure. As a result, Eq. (3) adopts a modified form:

$$\langle x_i x_j / J \rangle / \langle 1/J \rangle = kT(F^{-1})_{ij} \quad (3')$$

If the jacobian J is constant, Eq. (3′) reduces to Eq. (3). The functional form of J depends on the precise definitions of translational and rotational variables. Gonzalez and Maddocks show that for discretized continuum variables (motivated by the continuum theory of elastic rods), J is equal to one within a second-order correction in rotational variables. Lankas et al.[51] calculated the jacobian for *3DNA*, a popular algorithm for helicoidal analysis, and found that it is equal to one, on condition that the sum of squares of roll and tilt (in radians) is much less than one. In fact, J is always a nonzero constant at the absence of fluctuations, and changes only slightly if the fluctuations are small. Thus, for small deformations, the macroscopic and microscopic approach [Eq. (3) and (3′)] coincide, which could be expected.

A MODEL SYSTEM: *Eco*RI DODECAMER

As an example, we present the local (base-pair step) deformability calculated from MD simulations of a 16-bp DNA oligomer CGCGCGAATTCGCGCG containing *Eco*RI binding site. The central 12-mer, crystallized as early as in 1980,[53] has served as a model system in a number of DNA structural studies. Here, we performed a 18 ns MD (with explicit water and counterions in a rectangular box) using a protocol described elsewhere.[49,50] The *3DNA* analyzer was used to extract the time courses of conformational parameters from MD trajectories. Then we applied Eq. (3) to calculate the force constants with respect to the six base-pair step helicoidal parameters (twist, roll, tilt, shift, slide, rise), including all coupling terms. This was done for every base-pair step (except terminal ones). Thus, in addition to the usual profiles of average helicoidal parameters, one can now plot the "deformability profiles" along a sequence as well. Since our sequence is symmetric with respect to its center, so are expected to be the plots of the force constants.

Figure 1 shows the profiles of the "diagonal" force constants that reflect the free energy changes upon deforming only one conformational parameter (the others being kept at equilibrium values). Looking at twist, we indeed see an almost perfectly symmetric profile. Steps of the same composition exhibit similar deformability: the pyrimidine–purine (YR) steps CG are the most flexible, as contrasted with GC, and the central AT step (RY) is the stiffest. The data for roll show a lack of symmetry in the second GA (= TC) step. In fact, the whole second half of the oligomer exhibits higher flexibility, although the contrast between CG and GC steps and the high stiffness of the central AT step is preserved. This suggests possible different nonlocal substates in which the two halves are, which may be related to insufficient equilibration of ions. On the other hand, the profiles of stiffness in tilt, shift, slide, and especially rise in general show quite a good symmetry; nevertheless, the tendency of the second half being more flexible is apparent as well. Note the sharp contrast in deformability depending on the step composition and high deformability of YR steps in angular variables (twist, roll, tilt). The profile of rise shows that although symmetrically placed CG steps have very similar stiffness, not all CGs are identical: the outer ones are stiffer than the inner ones. This points to a possible influence of the surrounding sequence on the step deformability.

The code for constructing deformability profiles from MD is available on request from the author.

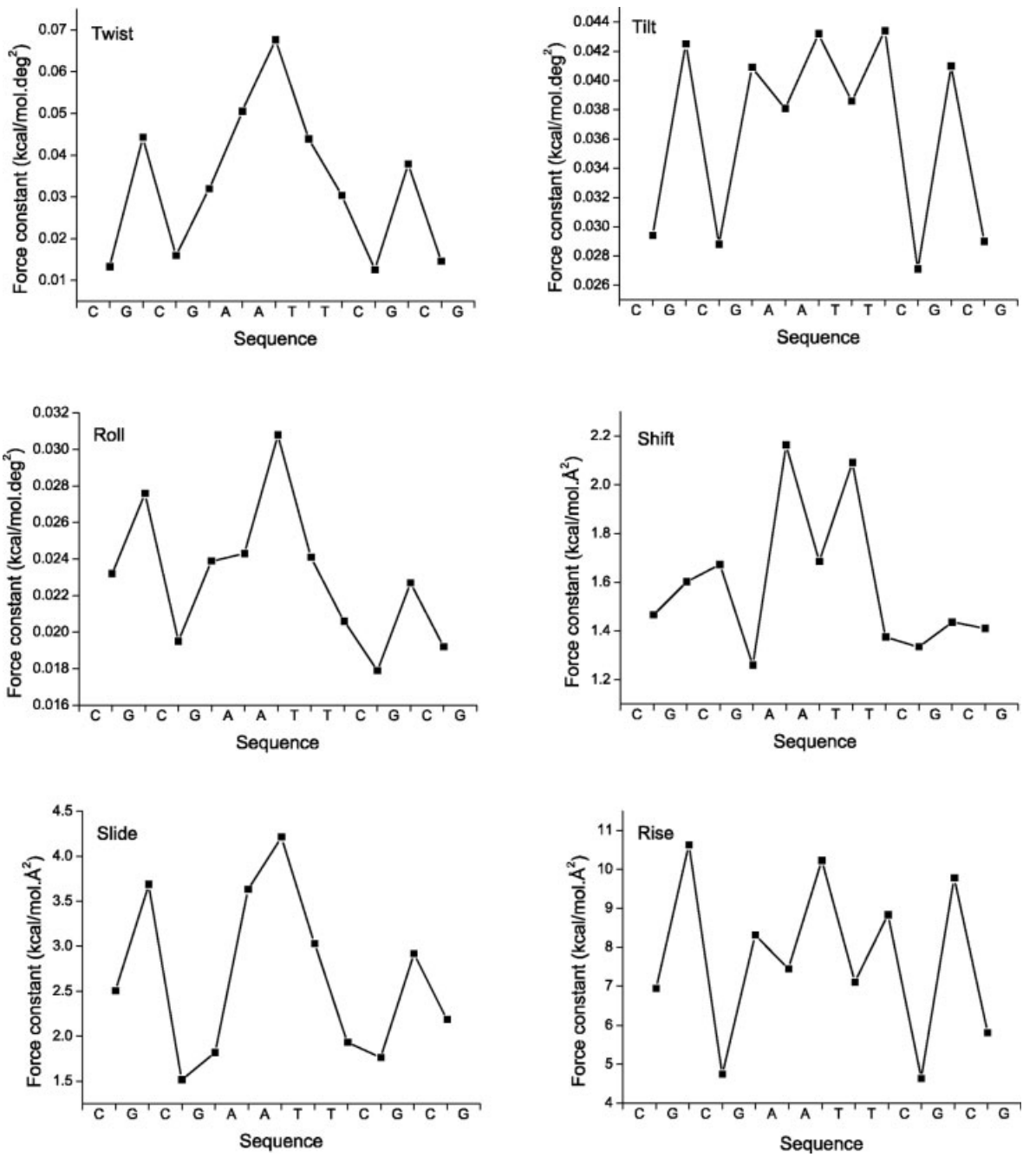

FIGURE 1 Deformability profiles of the local (base-pair step) force constants with respect to the six helicoidal variables for *Eco*RI dodecamer. The 12-bp sequence was flanked by CG steps at each end, which were not analyzed and are not shown in the picture. The results are based on an 18 ns atomic-resolution molecular dynamics simulation. The snapshots (taken every ps) were analyzed by the *3DNA* code to obtain time courses of the conformational parameters. Equation (3) was then used to calculate the force constants (including all coupling terms not shown here). While some of the profiles show a high degree of symmetry (twist, tilt, rise), others exhibit lower stiffness in the second part of the fragment (roll, shift). This may be related to trapping of the two halves in different substates, possibly due to insufficient ion equilibration.

COMPARING LOCAL FORCE FIELDS

As already mentioned above, several groups have developed local (base-pair step) force fields for DNA deformation using different methods. These force fields describe DNA deformability at various levels of detail and thus are not all directly comparable: the data of Olson and co-workers[28] and the MD data of Lankas et al.[51] contain the full stiffness matrix for the six conformational parameters, while other force fields usually report isotropic bending stiffness. Moreover, most force fields are relative and must be calibrated against a quantity known from elsewhere.

But some approximative comparison can be made. Here, we compare the force field of Robinson and co-workers based on an electron paramagnetic resonance technique,[23] the one by De Santis and co-workers inferred from thermal stability data,[25] the deformability parameters by Olson and co-workers (from an ensemble of crystal structures)[28] and our MD data.[51] The common parameter to compare is the relative isotropic bending rigidity. This quantity was constructed as follows. First, an isotropic bending rigidity for the data of Olson et al. and for MD data was calculated as harmonic average of tilt and roll stiffness. The other two force fields already contain this value. Then, the "generic" value was computed as harmonic average of the values of the ten unique steps. The result is then the value for a particular step divided by the generic value.

The comparison is shown in Figure 2a. Note that the first three steps on the x axis are of the YR type, the last three are RY, and the four in the middle are RR. The Robinson data show a general tendency of increasing stiffness from YR through RR to RY. The value for GC is not available, since the reciprocal value of the force constant was zero (within the sensitivity of the method); the most flexible and the stiffest steps thus differ by a factor of four or more. By contrast, the De Santis data show much smaller variations. The Olson parameters behave in a rather complicated way, changing between 0.8 and 1.6. The MD data exhibit a clear trend: they predict flexible YR steps, intermediate RR and stiff RY, spanning a range between 0.7 and 1.2. And so, the force fields differ both in the range of values and the trends they exhibit. Note, however, that the Olson and MD force fields are quite close to each other for YR and RY steps; for RR steps, the Olson data exhibit substantial variations while the MD values change much less.

It should be stressed that this comparison regards only relative bending stiffness and says nothing about the flexibility in twist or in translational parameters, since those are not available for all the four force fields. As we show elsewhere,[51] the trend in the MD data presented here in fact reflects the trend in roll stiffness, while tilt stiffness is low for YR and higher (and similar) for all the other steps. The same (a bit less pronounced) applies to twist. Shift and slide flexibilities show no clear trend,

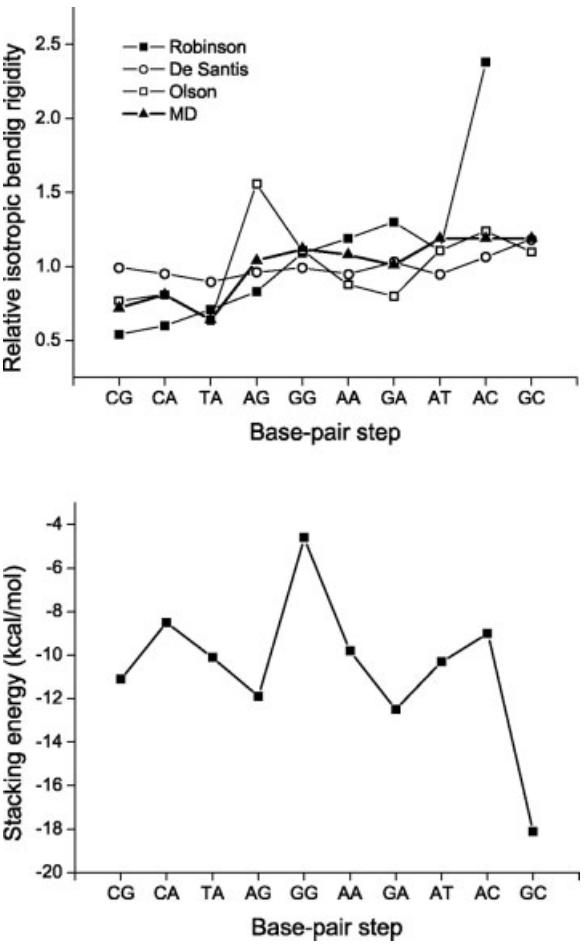

FIGURE 2 (a) Relative isotropic bending rigidity as a function of the dinucleotide sequence for local force fields inferred from an EPR resonance experiment (Robinson and co-workers), thermal stability data (De Santis and co-workers), ensemble of crystal structures (Olson and co-workers), and atomic resolution molecular dynamics simulations (Lankas et al.). The value for GC in the Robinson force field was not available since the inverse force constant was found to be zero within the sensitivity of the method. The Robinson force field exhibits a general increase in stiffness from pyrimidine–purine (YR) through RR to RY steps. The MD field predicts three distinct groups of steps: flexible YR, intermediate RR, and stiff RY. More detailed analysis shows that this trend is due to differences in roll deformability. Comparison with base-pair step stacking energies (b) from modern ab initio quantum chemical calculations (see text for details) shows that none of the force field is correlated with stacking energy values.

while rise stiffness generally increases from YR through RR to RY.

It is interesting to compare these data with other related works. In their MD study, McConnel and Beveridge[54] assigned a certain flexibility to each of the ten dinucleotide steps based on the area of their roll–tilt plots enclosing 98% of the MD data points. They found that YR steps are on average significantly more flexible than the others; the most flexible step was TA and the least flexible AT. Later, Thayer and Beveridge divided the roll–tilt plots into quadrants and 10 concentric rings, which enabled them to extract more detailed information about the distribution of MD data for use in their Hidden Markov model for DNA–protein interactions.[55]

In their studies mentioned above, Packer and Hunter calculated the stiffness with respect to slide in dinucleotides[34] and slide stiffness of the central step in tetranucleotides.[35] All the other conformational parameters are optimized. Their ratio of force constants between the stiffest and the most flexible step is more than 1:10. Among their findings one may notice that CG and GC exhibit similar flexibility on the dinucleotide level and that tetranucleotides with central AT step rank among the most flexible as well as the stiffest ones, depending on the base pairs flanking the central step.

An interesting question arises about a possible relationship between stiffness and stacking energy. The hypothesis that stacking energy may have a determining influence on DNA deformability was proposed by Hagerman.[56] De Santis and co-workers[24,25] report a very close correlation (0.96) between their stiffness scale and the stacking energies based on an early theoretical work.[57] However, the order of stability of the ten base-pair steps predicted by these old semiempirical calculations significantly differs from that predicted by the recent quantum chemical computations.[58,59] Modern electron correlation calculations represent almost converged estimates of stacking energies and provide reliable relative order of intrinsic stacking energy.[60] Comparing the data by De Santis and co-workers[25] with contemporary evaluation of base stacking in B-DNA geometries[58,59] leads to a correlation coefficient of only 0.3. Taking into account the quality of contemporary ab initio methods, we have to conclude that the thermal stability scale used by De Santis and co-workers is probably not correlated with intrinsic base stacking energies.

On the other hand, modern quantum chemical data are excellently reproduced by modern empirical force fields.[58,59] One might thus expect that the MD flexibility scale (calculated using the Cornell et al.[61] force field) will show better correlations. But as can be seen by comparing Figures 2a and 2b, it is by no means the case. In fact, the stacking energies shown in Figure 2b are not correlated with any of the force fields in Figure 2a. Thus, we suggest that the DNA stiffness may not be determined simply by the magnitude of base stacking but by a complex interplay of various factors such as potential energy profiles of the steps and balance between stacking and hydration effects.

GLOBAL DEFORMABILITY

DNA deformability on length scales much longer than one base-pair step (here denoted as "global") plays a role in many biological processes such as wrapping of DNA around nucleosomes or its packing in the nucleus. To connect this global scale with the local one, an appropriate coarse-graining scheme must be introduced. This task has been addressed by several authors.

First of all, this problem arises in various experimental techniques such as fluorescence polarization anisotropy,[16] electron paramagnetic resonance,[21–23] or transient polarization grating.[20] In these cases one wants to infer intrinsic DNA deformability from the signal of a probe that reflects global motion of the fragment. Song et al.[62] have proposed a semianalytical solution of the unconstrained Brownian dynamics for short fragments (up to about 60% of the persistence length). Their weakly bending rod model assumes that the fragment consists of identical, independent units with isotropic, harmonic bending potential. Later, Robinson and co-workers[22] extended the model so that the units (e.g., base-pair steps) can have different force constants, thus allowing for inhomogeneity of the fragment.

Local (base-pair step) deformability can also be related to various other macroscopic properties of long fragments, such as configurations corresponding to elastic energy minima. The model of Manning et al.[63] consists of identical units with harmonic bending and twisting potential (without coupling). Coleman et al.[64] developed a theory using units (not necessarily identical) whose deformation potentials are general functions of the six helicoidal parameters (twist, roll, tilt, shift, slide, rise). They investigate elastic equilibria of DNA minicircles with local deformation potentials being a simplified quadratic functions and show that the equilibrium shape can be substantially affected by taking the coupling terms in the deformation potentials into account. Zhang and Crothers[14] in their statistical mechanical model use independent units with uncoupled tilt, roll, and twist harmonic deformability, and relate the force constants to macroscopic

quantities such as the experimentally observable *J* factor.

A particularly interesting set of global parameters are continuum elastic constants of DNA. On long length scales, DNA can be described as a flexible rod with a certain stretch modulus and twist rigidity, with two (generally different) bending rigidities in two perpendicular directions, allowing for bending anisotropy and with all cross-terms. A specific problem now consists in connecting the local deformability to these global elastic properties. Schellman and Harvey[65] proposed a general scheme to infer both static and dynamic persistence length from local geometry and deformability. Manning et al.[63] took the global constants equal to the local ones (scaled by the number of base-pair steps), Coleman et al.[64] performed all calculations at the discretized level. Another approach has been adopted by De Santis and co-workers.[24,25,66] They assume that the relative rigidities of different base-pair steps (both with respect to bending and twisting) scale as relative melting temperatures of the steps. To calculate the global stiffness of a given sequence, they compute the arithmetic average of the relative melting temperatures of all steps in the sequence, by which they multiply a generic persistence length to get the actual persistence length of that particular sequence. Thus, their stiffness is a position-independent parameter of the sequence and can be put in front of any integrals along the sequence. This assumption enables them to solve analytically problems such as computing free energy changes upon nucleosome formation.

Another approach has been proposed by Matsumoto and Go.[67] They computed normal modes of a DNA fragment using a local force field and compared them to the normal modes of a continuum isotropic flexible rod. The latter depend on the elastic constants of the rod and can be calculated analytically. Thus, by fitting the normal modes of the two systems, the elastic constants of the rod can be estimated. While establishing the twisting and bending rigidity, they were unable to calculate the stretch modulus, as no observed mode corresponded to pure stretching. This limitation has been overcome in a later work by Matsumoto and Olson[68] where a base-pair oriented coordinate system together with a previously developed dinucleotide force field[28] was used. The authors show that the three lowest modes of bending, twisting as well as stretching indeed scale with the fragment length (between 40 and 200 bp) almost exactly as the modes of a homogeneous isotropic elastic rod, so that the elastic constants can be readily computed.

A certain limitation of the above-mentioned approaches is that they a priori assume specific form of the local (e.g., base-pair step) force field. One either fixes its functional form and fits the parameters in it using macroscopic experiments (such as fluorescence or EPR techniques, or cyclization), or one infers some global properties (minicircle shape, continuum elastic constants) starting from the local field with given parameters.

This limitation can be circumvented by defining global variables to describe the deformation of the fragment as a whole. The widely used conformational analyzer *Curves* quantifies both direction and magnitude of global bending and this feature has been used to analyze global properties of A-tracts simulated by MD.[69] An alternative concept is represented by the *Madbend* algorithm proposed by Schlick and co-workers[70]: they define two global bending angles (global roll and global tilt) that are, in principle, sums of base-pair step rolls and tilts compensated for twist. Schlick and co-workers later combined this approach with principal component analysis to investigate global motions as part of their study of 13 different TATA sequences.[71]

Some time ago, Lankas et al.[49] proposed a simple method for extracting global elastic properties directly from atomic-resolution MD of DNA oligomers, without any assumptions regarding the local (base-pair step) interactions. They first define global deformation variables of the fragment as a whole: total fragment length, total twist, and two bending angles (towards the grooves and towards the backbone in the fragment center). The fluctuations of these variables recorded along a MD trajectory are inserted into eq. (3) to give the fragment's stretch modulus, twist rigidity, and two bending rigidities in two perpendicular bending directions. All the coupling terms in the elastic energy expression (such as twist–stretch coupling) can be calculated as well. The advantage of this approach is that possible nonlocal effects (shorter than the fragment length) are automatically taken into account. The authors studied four 17-bp oligomers (11-bp stretches of A, G, AT, and CG, capped with GCG at each end); the trajectory length was 5 ns. The GCG caps were excluded from the analysis. The resulting elastic constants were in a very good overall agreement with experimental values for random sequences. The atomic-resolution MD approach, however, revealed pronounced sequence dependence of the stretch modulus and torsional rigidity, while bending rigidity differed only slightly among the sequences studied. Only polyAT was revealed as unambiguously more flexible in bending compared to other sequences. The exceptional flexibility of AT-tracts is supported by recent experimental results: Robinson and co-workers[22] in their EPR study found AT- tracts

to be 20% more flexible than the control sequence; Zhang and Crothers[15] measured a similar decrease ($28 \pm 12\%$) by cyclization kinetics. On the other hand, Matsumoto and Olson,[68] who studied long stretches of A, G, AT, and GC by normal mode analysis, do not confirm this trend.

The value of twist–stretch coupling inferred from MD was in quantitative agreement with experiments as well (see Lankas et al.[49] for further discussion). The twist–bend coupling, theoretically predicted earlier,[72] was calculated for the first time and found out to be the most important cross-term for fragments shorter than one helical turn.

Later, Lankas et al.[50] used the same method as part of their study on the effect of the N2 amino group on structure, dynamics, and elasticity of polypurine tracts. They again simulated 17-bp oligomers with central 11-bp parts formed by stretches of A, I (inosine—paired with 5-methylcytosine), AI, G, and D (2-aminoadenine). A-tract, I-tract, and AI-tract were found to have high stretch modulus and low twist rigidity, as contrasted to G-tract and D-tract with low stretch modulus and high twist stiffness. This supports the structural findings in the study—namely, that the presence or absence of the minor groove amino group rather than the electrostatic similarity of the base pairs decides the properties of the tracts. The differences in bending stiffness among the tracts were found to be rather small.

To investigate a possible dependence of the results on the time scale of the simulations, the trajectories for A-, G-, and I-tract were prolonged from 5 to 20 ns. The average conformational parameters (from the average structure) differ in most cases by only several percent compared to those from the shorter simulations. The elastic constants though, while fully maintaining the observed trends, somehow differ. This regards especially the twist rigidity of the A-tract, which was exceptionally low in the shorter simulations (5 ns) but which now (after 20 ns) adopted a value closer to those of the other sequences (but still relatively low). This suggests that while the average structure is rather robust and remains quite similar after several ns of simulations, elastic properties are much more sensitive.

Another point of concern is the precise definition of the global deformation variables. A similar problem has been encountered on the local level and extensively analyzed.[73,74] Lankas et al.[49] defined their global variables by means of two helicoidal analysis algorithms: *Curves*[75,76] and *CEHS*.[77] The *Curves* code has the unique property to construct a global, curvilinear helical axis of the fragment—this axis was used to define the total fragment length while its tangent vectors defined the base-pair normals. On the other hand, *CEHS* is a purely local scheme and the fragment length in this case was defined simply as the sum of distances of the successive base-pair centers. The resulting elastic constants in the two cases are rather different, the *Curves* values being closer to the results of macroscopic experiments (interestingly, the exceptional bendability of the AT-tract was reproduced in both cases).

Table I Selected Global Elastic Constants of the Simulated *Eco*RI Dodecamer

	3DNA	Curves
Twist persistence length (nm)	77	65
Stretch modulus (pN)	2114	1752
Twist–stretch coupling (nm)	24	4
Isotropic bending persistence length (nm)	69	68

One may argue that *Curves* and *CEHS* schemes differ a lot also on the local level, without making use of the curvilinear axis.[73] To investigate the point further, we report here global elastic constants of our model system, the *Eco*RI dodecamer (see Table I) analyzed by *Curves* (using the global curvilinear axis) and by *3DNA*,[78] a popular analyzer that, on the local level, falls into the same category as *Curves*.[73] Only the central 12 bp of the 16-bp fragment were analyzed. As can be seen in Table I, the local scheme produces somehow higher twist rigidity, higher stretch modulus, as well as higher twist–stretch coupling. The isotropic bending stiffness differs only slightly. The same trends were observed previously with *CEHS*, but were much more pronounced.

The code for extracting global elastic constants from MD simulations using both *3DNA* and *Curves* is available on request from the author.

On the base-pair step level, DNA bending is anisotropic: it is much easier to bend DNA into grooves (by changing roll) than to backbones (tilt). This has also been found in the early theoretical study of Zhurkin et al.[32] Lankas et al.[49] studied the dependence of the elastic constants on the fragment length (for 3–11-bp fragments), and found that bending anisotropy, clearly pronounced for trimers, disappears for fragments of approximately half the repeat length and reappear for longer ones, yet to a much lesser extent. This suggests that for DNA fragments of several helical turns or more, the bending anisotropy will generally not play any role. Moreover, while the two "true" bending rigidities differ quite a lot depending on the fragment length, their harmonic average (referred to as isotropic bending rigidity) remains almost constant. This is

in line with theoretical findings by Kehrbaum and Maddocks,[79] who showed that an elastic rod with locally anisotropic bending, but a high intrinsic twist, is well approximated on relatively long length scales by an effective rod with an isotropic bending law, the effective isotropic bending constant being the harmonic average of the two local ones. However, as Matsumoto and Olson[68] point out, insertion of phased fragments with exceptional bending properties can lead to anisotropic effects on mesoscopic scale, just as the phased A-tracts give rise to the global curvature.

An open problem concerns the possible influence of nonlocal effects on DNA global properties. To which extent can DNA be described by nearest-neighbor interactions? Specifically, can DNA elastic properties at longer length scales be inferred from local (base-pair step) interactions? This is an underlying assumption of all the base-pair step force fields. But Robinson and co-workers[23] found that to construct their force field with all 10 base-pair steps considered different, 9 out of 40 used sequences must be discarded—they do not satisfactorily obey the dinucleotide model. In their theoretical studies, Packer and Hunter[34,35] suggest that while the structure of some dinucleotides is context independent, there are cases where the influence of context is important, and the agreement with crystallographic data dramatically improves upon transition to tetranucleotides. The dependence of global elastic constants on the fragment length calculated by Lankas et al.[49] exhibits an initial increase on passing from trimers to 4–5-bp fragments. Correlation coefficients of helicoidal parameters along a sequence, which we report elsewhere,[51] unambiguously show that base-pair steps do not behave as independent units but their movements are correlated. The correlation typically extends to second or third nearest neighbor.

CONCLUSIONS AND PERSPECTIVES

Recent developments in experimental and computational techniques make it possible to start a systematic study of DNA sequence-dependent deformability. The data based on an ensemble of crystal structures[28] can be further refined and possibly also extended beyond the dinucleotide approximation as new structures are accumulating. We can expect further developments in the field of fluorescent and EPR techniques, which have shown their power in a number of results including a dinucleotide force field.[23] Latest developments of the cyclization technique[14,15] pave the way for high-throughput investigation of DNA mechanical properties at equilibrium and in more physiological conditions.

As the present article attempts to show, modern computational methods may represent an indispensable source of information in this context. Atomic resolution molecular dynamics simulations using up-to-date empirical force fields and accurate treatment of long-range electrostatic interactions have proved their validity in a number of studies on proteins and nucleic acids. They are now able to deal with system sizes and time scales on which a relevant information about DNA deformability can be extracted. The unique feature of these methods is that one can trace down each and every atom and thus investigate the microscopic origin of the system's macroscopic behavior. With further development of computer technology and simulation algorithms, the possibilities of this approach will certainly broaden.

However, producing longer trajectories on larger systems is not enough. It must be accompanied by further development of theoretical concepts to extract useful information from the simulations. One of the important topics are theoretically well founded and consistent coarse-graining schemes. Within the framework of elasticity theory, one needs constitutive relationships for elastic bodies such as rods that would include nonlocal effects observed in both experiment and simulation. The fluctuations produced due to nonzero temperature of the simulated system, used in a limited way up to now, represent a source of information that is still awaiting a full exploitation. This all will contribute to better understanding of DNA properties and its biological function.

The author wishes to thank Jiri Sponer, Thomas E. Cheatham III, John H. Maddocks, and Jörg Langowski for valuable discussions. This work was partially supported by the project "Center for Complex Molecular Systems and Biomolecules" (LN00A032) financed by the Ministry of Education of the Czech Republic, by the project "Properties of DNA on Different Length Scales" provided by the Volkswagen Foundation, and by NATO grant no. LST.CLG.977846.

REFERENCES

1. Olson, W. K.; Zhurkin, V. B. Curr Opin Struct Biol 2000, 10, 286-297.
2. Smith, S. B.; Finzi, L.; Bustamante, C. Science 1992, 258, 1122-1126.
3. Smith, S. B.; Cui, Y.; Bustamante, C. Science 1996, 271, 795-798.
4. Bustamante, C.; Marko, J. F.; Siggia, E. D.; Smith, S. Science 1994, 265, 1599-1600.

5. Cluzel, P.; Lebrun, A.; Heller, C.; Lavery, R.; Viovy, J.-L.; Chatenay, D.; Caron, F. Science 1996, 271, 792-794.
6. Strick, T. R.; Allemand, J.-F.; Bensimon, D.; Bensimon, A.; Croquette., V. Science 1996, 271, 1835-1837.
7. Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. Proc Natl Acad Sci USA 1997, 94, 6185-6190.
8. Landau, L. D.; Lifshitz, E. M. Theory of Elasticity, 3rd ed.; Butterworth-Heinemann: Oxford, UK, 1986.
9. Shore, D.; Langowski, J.; Baldwin, R. L. Proc Natl Acad Sci USA 1981, 78, 4833-4837.
10. Shore, D.; Baldwin, R. L. J Mol Biol 1983, 170, 957-981.
11. Roychoudhury, M.; Sitlani, A.; Lapham, J.; Crothers, D. M. Proc Natl Acad Sci USA 2000, 97, 13608-13613.
12. Widlund, H. R.; Cao, H.; Simonsson, S.; Mangusson, E.; Simonsson, T.; Nielsen, P. E.; Kahn, J. D.; Crothers, D. M.; Kubista, M. J Mol Biol 1997, 267, 807-817.
13. Nathan, D.; Crothers, D. M. J Mol Biol 2002, 316, 7-17.
14. Zhang, Y.; Crothers, D. M. Biophys J 2003, 84, 136-153.
15. Zhang, Y.; Crothers, D. M. Proc Natl Acad Sci USA 2003, 100, 3161-3166.
16. Allison, S. A.; Schurr, J. M. Chem Phys 1979, 41, 35-39.
17. Millar, D. P.; Robbins, R. J.; Zwail, A. H. J Chem Phys 1982, 76, 2080-2093.
18. Schurr, J. M. Chem Phys 1984, 84, 71-76.
19. Pedone, F.; Mazzei, F.; Matzeu, M.; Barone, F. Biophys Chem 2001, 94, 175-184.
20. Naimushin, A. N.; Fujimoto, B. S.; Schurr, J. M. Biophys J 2000, 78, 1498-1518.
21. Okonogi, T. M.; Reese, A. W.; Alley, S. C.; Hopkins, P. B.; Robinson, B. H. Biophys J 1999, 77, 3256-3276.
22. Okonogi, T. M.; Alley, S. C.; Reese, A. W.; Hopkins, P. B.; Robinson, B. H. Biophys J 2000, 78, 2560-2571.
23. Okonogi, T. M.; Alley, S. C.; Reese, A. W.; Hopkins, P. B.; Robinson, B. H. Biophys J 2002, 83, 3446-3459.
24. Anselmi, C.; Santis, P. D.; Paparcone, R.; Savino, M.; Scipioni, A. Biophys Chem 2002, 95, 23-47.
25. Scipioni, A.; Anselmi, C.; Zuccheri, G.; Samori, B.; De Santis, P. Biophys J 2002, 83, 2408-2418.
26. Gotoh, O.; Tagashira, Y. Biopolymers 1981, 20, 1033-1042.
27. Landau, L. D.; Lifshitz, E. M. Statistical Physics, Part 1; 3rd ed.; Butterworth-Heinemann: Oxford, UK, 1980.
28. Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. Proc Natl Acad Sci USA 1998, 95, 11163-11168.
29. Isaacs, R. J.; Spielmann, H. P. J Mol Biol 2001, 307, 525-540.
30. McAteer, K.; Kennedy, M. A. J Biomol Struct Dyn. 2000, 17, 1001-1009.
31. Warmlander, S.; Sponer, J. E.; Sponer, J.; Leijon, M. J Biol Chem 2002, 277, 28491-28497.
32. Zhurkin, V. B.; Lysov, Y. P.; Ivanov, V. I. Nucleic Acids Res 1979, 27, 1081-1096.
33. Lafontaine, I.; Lavery, R. Curr Opin Struct Biol 1999, 9, 170-176.
34. Packer, M. J.; Dauncey, M. P.; Hunter, C. A. J Mol Biol 2000, 295, 71-83.
35. Packer, M. J.; Dauncey, M. P.; Hunter, C. A. J Mol Biol 2000, 295, 85-103.
36. Packer, M. J.; Hunter, C. A. J Am Chem Soc 2001, 123, 7399-7406.
37. Beveridge, D. L.; McConnell, K. J. Curr Opin Struct Biol 2000, 10, 182-196.
38. Cheatham, T. E., III; Young, M. A. Nucleic Acids Sci (Biopolymers) 2001, 56, 232-256.
39. Karplus, M.; McCammon, J. A. Nature Struct Biol 2002, 9, 646-652.
40. Boczko, E. M.; Brooks, C. L., III. J Phys Chem 1993, 97, 4509-4513.
41. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. J Comp Chem 1992, 13, 1011-1021.
42. Varnai, P.; Lavery, R. J Am Chem Soc 2002, 124, 7272-7273.
43. Banavali, N. K.; MacKerell, A. D., Jr. J Mol Biol 2002, 319, 141-160.
44. Huang, N.; Banavali, N. K.; MacKerell, A. D., Jr. Proc Natl Acad Sci USA 2003, 100, 68-73.
45. Varnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B. Nucleic Acids Res 2002, 30, 5398-5406.
46. Fuxreiter, M.; Luo, N.; Jedlovszky, P.; Simon, I.; Osman, R. J Mol Biol 2002, 323, 823-834.
47. Seibert, E.; Ross, J. B. A.; Osman, R. Biochemistry 2002, 41, 10976-10984.
48. Bruant, N.; Flatters, D.; Lavery, R.; Genest, D. Biophys J 1999, 77, 2366-2376.
49. Lankas, F.; Sponer, J.; Hobza, P.; Langowski, J. J Mol Biol 2000, 299, 695-709.
50. Lankas, F.; Cheatham, T.E., III.; Spackova, N.; Hobza, P.; Langowski, J.; Sponer, J. Biophys J 2002, 82, 2592-2609.
51. Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T.E., III. Biophys J 2003, 85, 2872–2883.
52. Gonzalez, O.; Maddocks, J. H. Theor Chem Acc 2001, 106, 76-82.
53. Wing, R. M.; Drew, H. P.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Nature 1980, 287, 755-758.
54. McConnell, K. J.; Beveridge, D. L. J Mol Biol 2001, 314, 23-40.
55. Thayer, K. M.; Beveridge, D. L. Proc Natl Acad Sci USA 2002, 99, 8642-8647.
56. Hagerman, P. J. Ann Rev Biophys Biophys Chem 1988, 17, 265-286.
57. Ornstein, R. L.; Rein, R.; Breen, D. L.; Macleroy, R. D. Biopolymers 1978, 17, 2341-2360.
58. Sponer, J.; Florian, J.; Ng, H.-L.; Sponer, J. E.; Spackova, N. Nucleic Acids Res 2000, 28, 4893-4902.

59. Sponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. Biophys J 1997, 73, 76-87.
60. Hobza, P.; Sponer, J. J Am Chem Soc 2002, 124, 11802-11808.
61. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179-5197.
62. Song, L.; Allison, S. A.; Schurr, J. M. Biopolymers 1990, 29, 1773-1791.
63. Manning, R. S.; Maddocks, J. H.; Kahn, J. D. J Chem Phys 1996, 105, 5626-5646.
64. Coleman, B. D.; Olson, W. K.; Swigon, D. J Chem Phys 2003, 118, 7127-7140.
65. Schellman, J. A.; Harvey, S. C. Biophys Chem 1995, 55, 95-114.
66. Anselmi, C.; Bocchinfuso, G.; Santis, P. D.; Savino, M.; Scipioni, A. Biophys J 2000, 79, 601-613.
67. Matsumoto, A.; Go, N. J Chem Phys 1999, 110, 11070-11075.
68. Matsumoto, A.; Olson, W. K. Biophys J 2002, 83, 22-41.
69. Sprous, D.; Young, M. A.; Beveridge, D. L. J Mol Biol 1999, 285, 1623-1632.
70. Strahs, D.; Schlick, T. J Mol Biol 2000, 301, 643-663.
71. Qian, X.; Strahs, D.; Schlick, T. J Mol Biol 2001, 308, 681-703.
72. Marko, J. F.; Siggia, E. D. Macromolecules 1994, 27, 981-998.
73. Lu, X.-J.; Olson, W. K. J Mol Biol 1999, 285, 1563-1575.
74. Lu, X.-J.; Babcock, M. S.; Olson, W. K. J Biomol Struct Dyn. 1999, 16, 833-843.
75. Lavery, R.; Sklenar, H. J Biomol Struct Dyn 1988, 6, 63-91.
76. Lavery, R.; Sklenar, H. J Biomol Struct Dyn 1989, 6, 655-667.
77. Lu, X.-J.; El Hassan, M. A.; Hunter, C. A. J Mol Biol 1997, 273, 668-680.
78. Lu, X.-J.; Shakked, Z.; Olson, W. K. J Mol Biol 2000, 300, 819-840.
79. Kehrbaum, S.; Maddocks, J. H. In 16th IMACS World Congress: Lausanne, Switzerland, 2000.

Reviewing Editor: Dr. David A. Case