9. Summary

This PhD thesis is focused on the first applications of a Monte Carlo (MC) algorithm for nucleic acids defined in the space of collective and internal variables. First results on sequence-specific conformation and flexibility of nucleic acid decamers and the medically important methylene blue (MB)-DNA complexes are discussed. The performance and efficiency of the MC algorithm was investigated by analyzing the equilibration of the systems and the effects of the starting configurations on the simulation results. The MC studies are complemented by energy minimizations of MB-DNA complexes and by evaluating the energy of these complexes using a continuum treatment of solvent electrostatic effects.

The MC algorithm described in chapter 4 is based on a molecular model with rigid bases and fixed bond lenghts resulting in a description of the molecular geometry by a set of 12 collective and internal variables per nucleotide. The nucleotides are moved by collective variables based on three rigid body translations and three rigid body rotations. The furanose puckering is varied by the MC variables sugar phase and amplitude following the pseudorotation concept of sugar puckering description [19, 152]. The glycosidic angle χ, the endocyclic dihedrals ε and γ and the valence angle ω_1 complete the set of 12 MC variables per nucleotide which are necessary for defining uniquely the conformation of each nucleotide. The MC moves of these variables are combined with an analytical chain closure procedure varying the other endocyclic backbone torsion and valence angles as dependent variables. In this way the conformational changes become local and the moves are restricted to single nucleotides. The MC acceptance criterion follows the Metropolis algorithm [143] and includes associated Jacobians [150]. For energy calculations the AMBER force field [106] and an implicit electrostatic solvent description [110] are used. The polyanionic charge of nucleic acids is neutralized by explicit counterions which are moved by additional MC variables describing ion translations.

Results of MC simulations of the DNA decamers $(\text{dA-dT})_5]_2$ und $[(\text{dG-dC})_5]_2$ are discussed in chapter 5. Equilibrations of the MC simulations are assessed on the basis of averaged energies and structural parameters. Equilibration is reached when the averages of the total energy as well as of the parameters such as x-dislacement, rise, inclination, helix twist, sugar phase and glycosidic angle remain approximately constant after $4 \cdot 10^5$ macrocycles indicating an equilibration. Structural parameters of nucleotides at the outer edges of the decamers indicate a higher flexibility resulting in a less perfect equilibration. The decamer with alternating AT base sequence also shows a higher flexibility resulting in a less perfect equilibration in the comparison of the two decamers with AT and GC alternating base sequences. This different flexibility of the two DNA decamers is based on reduced hydrogen bonding interactions stabilizing AT base pairs compared to GC base pairs.

The MC algorithm allows equilibrations of MC simulations on reasonable computational time scales estimated in chapter 4 whereas equilibration in MD simulations cannot be assessed [55, 64, 72]. In the MC sampling, transitions between substates are seen frequently in contrast to MD simulations [55, 72]. DNA conformations move from B-DNA toward both A- and D-DNA states but always return back to B-DNA. The palindromic symmetry of the base sequences of each decamer as reflected by the average values of the structural parameters and their fluctuations is used as a criterion for equilibration. In chapter 5 it is shown that the results of MC simulations do not depend on starting configuration. Average structures resulting from two equilibrated MC simulations, one starting from a B-DNA conformation and the other from an extremely deformed structure, have been calculated. An rmsd value of 0.20 Å between the two structures indicated that the two confirmations were essentially identical.
The average values and fluctuations of structural parameters show a different sequence specificity. The distinct specificity of the parameters including x-displacement, rise, tip, helix twist, sugar phase and glycosidic angle are discussed for the two DNA decamers with alternating AT and GC base sequences, respectively. The average values of the sugar pucker amplitude and of the dihedrals ε, ζ, α, γ, and δ as well as the fluctuations of the parameters rise, inclination, tip, helix twist, sugar phase, amplitude, glycosidic angle, and of the dihedrals α, β, γ und δ are found to be clearly sequence-specific, and are in good agreement with experimentally determined values [39, 45]. The important parameter helix twist in the range of 32°-34° for ApT, 35°-37° for TpA, 35.5°-36.5° for GpC, and 32.3°- 33.5° for CpG base-pair steps corresponds very well with experimental data [39] showing higher twists for YpR compared to RpY base steps in AT base sequences and conversely higher twists for RpY compared to YpR base steps in GC base sequences. In contrast to the present MC results, the helix twist values calculated from MD simulations are around 30° and show a significant underwinding by 3-4° compared to experimental data [55, 64, 72].
The movements of the explicit counterions have been also analyzed statistically. The higher occupancy of ions in the surrounding of the decamers fits qualitatively with the expected counterion condensation at the DNA surface [47]. Defining a condensation radius of 22 Å around the helical axis, ca. 61 % of the counterions condensate at the DNA surface in accordance with experimental and theoretical studies [48, 161].
Results on the medically important methylene blue (MB)-DNA binding obtained by energy minimizations are discussed in chapter 6. Lowest energy MB-DNA complex structures representing the three different binding modes (intercalation between neighbouring base pairs, binding of MB by insertion into the minor or the major groove) were derived [53]. The MB-DNA complexes were selected following the criterion of lowest total energies taking into account electrostatic solvent effects by a continuum treatment using the FDPB method [114, 115]. MB intercalates into the 5'-TpA-3' and the 5'-ApT-3' base steps of the AT alternating decamer as well as into the 5'-CpG-3' and the 5'-GpC-3' base steps of the GC alternating decamer. An adiabatic mapping by rotating the long axis of MB about the helical axis results for each intercalation site in two different MB orientations. The symmetric intercalation is characterized by an orientation of the long axis of MB parallel to the flanking base pairs whereas the gauche intercalation refers to a rotation of the long axis of MB around the helical axis by approximately 140°-145°. Symmetric intercalation has been found to be energetically favored in a salt free environment. The estimated binding energies of both MB-DNA

complexes with symmetric intercalation of MB at the 5'-CpG-3' and the 5'-GpC-3' site indicate equal stabilities in agreement with spectroscopic data [100]. The lower stability of the gauche intercalation structures compared to the symmetric orientation of MB is due to reduced stacking interactions of the dye with the flanking base pairs and the larger deformation of the target decamers.
Salt effects on the MB-DNA binding have been studied by estimating binding energies as a function of monovalent salt concentration in the range of $10^{-4} - 2$ mol/l. The minor groove binding of MB to the AT alternating decamer is stabilized at low ionic strength and remains energetically favored also at high ion concentrations. The intercalative binding of MB to the GC alternating decamer that is energetically favored at low ionic strength changes to minor groove binding at high salt concentrations [76]. The results on sequence specificity and salt dependence of MB-DNA binding are in agreement with experimental data [100, 173]. The predicted structures of MB-DNA complexes presented in chapter 6 could serve as a basis for interpretation of experimental studies since neither a high-resolution structure derived from X-ray or NMR data is yet available nor the architecture of the complexes can be precisely determined by spectroscopic data [177, 178].
The results of the MC simulations on MB-DNA complexes, allows for the first time a thorough understanding of the dynamical properties of DNA-ligand binding. MD simulations are restricted to small systems [184] or time scales that do not allow to reach equilibration in the calculations [84]. In chapter 7 it is shown that the current limitations of MD simulations for studying dynamic properties of DNA-ligand complex formations [185] are overcome by using the MC algorithm described in chapter 4. In addition to the degrees of freedom of flexible DNA, the MC algorithm uses three translations and three rotations for describing the ligand movement relative to the target DNA. These six variables are complemented by four internal MC variables of methyl group rotations.
In addition, the results of MC simulations show that both the symmetric and the gauche intercalations are stable conformations as transitions between the two structural alternatives do not occur. The position and orientation of MB in the intercalation pocket are less variable in the case of a symmetric intercalation compared to a gauche intercalation. The symmetric intercalation at the 5'-GpC-3' step is an exception since the long axis of MB rotates around the helical axis by approximately 60° but returns to the symmetric intercalation with an average rotation of 1°-5°. The conformation of the target decamers shows a stretching and unwinding at the intercalation pocket. Higher rise values at gauche intercalations indicate a larger deformation of the target decamers in comparison to symmetric intercalations. The differences between average helix twist values of YpR and RpY base-pair steps of free decamer structures are maintained in deformed intercalation complexes. Conserved helix twist differences have been observed at underwound base-pair steps forming an intercalation pocket as well as at overwound base-pair steps adjacent to the intercalation pockets which counterbalance the underwinding at the binding pocket. Sugar puckering modes of the nucleotides flanking the intercalation site are shown to be sequence-specific as they differ for intercalations at YpR and RpY sites.
Minor groove binding of MB to the AT alternating decamer show frequent transitions between alternative binding sites. The discrete binding sites of MB are characterized by the methyl groups facing outside the minor groove and by the sulfur and the central

nitrogen atom of MB lying approximately within a base-pair plane. Minor groove binding of MB to the GC alternating decamer shows only few sporadic transitions between alternative binding sites. This higher localisation of MB within the minor groove is due to a hydrogen bond formed between the central nitrogen atom of MB and a proton of the amino group of a guanine. Binding of MB in the major groove shows much larger movements of the ligand indicating a lower binding stability. Groove binding of MB is associated with a bending of the target decamer going along with decreased averaged rise values of all base-pair steps and increased average helix twist values of YpR base-pair steps.

The MC simulations of MB-DNA complexes identify stable binding states. They describe transitions between alternative binding sites of MB and sequence-specific deformations of the target nucleic acids upon MB binding. The results on the flexibility of MB-DNA complexes contribute to a deeper understanding of ligand binding and could assist in detailed interpretations of experimental studies.

It is expected that the new MC approach will be applied to a wide range of biologically important systems. The current studies will be extended to MC simulations of additional sequence-specific nucleic acid structures as well as of sequence-specific protein-DNA interactions.