



# HHS Public Access

Author manuscript

*J Comput Chem.* Author manuscript; available in PMC 2019 April 29.

Published in final edited form as:

*J Comput Chem.* 2018 October 30; 39(28): 2360–2370. doi:10.1002/jcc.25571.

## A new protein nucleic-acid coarse-grained force field based on the UNRES and NARES-2P force fields

Adam K. Sieradzan<sup>1</sup>, Artur Giełdo <sup>1</sup>, Yanping Yin<sup>2</sup>, Yi He<sup>2,3</sup>, Harold A. Scheraga<sup>2</sup>, and Adam Liwo<sup>1,\*</sup>

<sup>1</sup>Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland <sup>2</sup>Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, 14853-1301, U.S.A. <sup>3</sup>School of Engineering, University of California Merced, 5200 N. Lake Road, Merced, CA 95343, U.S.A

### Abstract

Based on the coarse-grained UNRES and NARES-2P models of proteins and nucleic acids, respectively, developed in our laboratory, in this work we have developed a coarse-grained model of systems containing proteins and nucleic acids. The UNRES and NARES-2P effective energy functions have been applied to the protein and nucleic-acid components of a system, respectively, while protein – nucleic-acid interactions have been described by the respective coarse-grained potentials developed in our recent work [Yin et al., *J. Chem Theory Comput.*, **2015**, 11, 1792–1808]. The Debye-Hückel screening has been applied to the electrostatic-interaction energy between the phosphate groups and charged amino-acid side chains. The model has been integrated into the UNRES package for coarse-grained molecular dynamics simulations of proteins and the implementation has been tested for energy conservation in microcanonical molecular dynamics runs and for temperature conservation in canonical molecular dynamics runs. Two case studies were performed: (i) the dynamics of the Ku protein heterodimer bound to DNA, for which it was found that the Ku70/Ku80 protein complex plays an active role in DNA repairing and (ii) conformational changes of the multiple antibiotic resistance (MarA) protein occurring during DNA binding, for which the functionally-important motions occurring during this process were identified.

### Graphical Abstract

\*Corresponding author; phone: +48585235350; fax: +48585235350; adam@sun1.chem.univ.gda.pl.

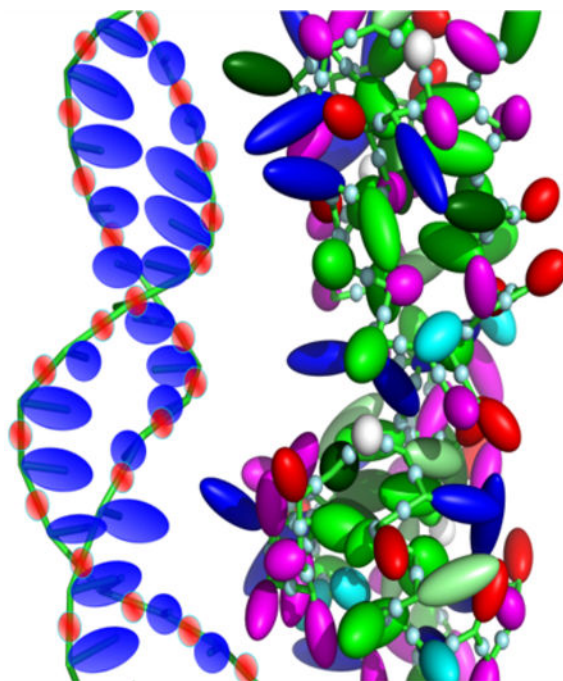


Illustration of the coarse-grained representation of protein – nucleic-acid systems implemented in this work with the example of the multiple antibiotic resistance (MarA) protein – DNA complex. Left: the DNA fragment in NARES-2P representation, right: the protein fragment in UNRES representation. The virtual bonds are shown as sticks, united side chains and nucleic-acid bases are shown as ellipsoids of revolution, united peptide groups are shown as light-blue spheres and united phosphate groups are shown as light-red spheres.

### Keywords

Protein-nucleic acid interactions; Coarse graining; UNRES; NARES-2P; Molecular dynamics

## INTRODUCTION

Modeling the structure and dynamics of biological macromolecules and molecular assemblies is nowadays of utmost importance in the quest for the understanding of the functioning of the machinery of life. Owing to the use of world-distributed computing (the FOLDING@HOME project)<sup>1</sup> the development of very efficient load-balanced parallel codes such as GROMACS<sup>2</sup>, NAMD<sup>3</sup>, or DESMOND<sup>4</sup>, the implementation of all-atom molecular dynamics (MD) programs on graphical processor units (GPUs)<sup>5</sup>, and the construction of dedicated machines<sup>6,7</sup>, enormous progress has been achieved in the field of all-atom molecular simulations in the last two decades. Nevertheless, the biological size- and time-scale is still beyond reach with the all-atom approach.

Coarse graining, which is a procedure to reduce the representation of a system under study from atomistic-detailed to a less detailed one, provides an intermediate resolution at which it is possible to handle large systems<sup>8-11</sup>. In the last two decades, we have been developing the

physics-based UNited RESidue (UNRES) model for the simulations of the structure and dynamics of proteins<sup>12,13</sup> and, quite recently, the NARES-2P model of nucleic acids<sup>13–15</sup>. The effective energy functions of UNRES and NARES-2P originate from the potential of mean force of the system under study, in which the degrees of freedom not present in the model (secondary degrees of freedom) are integrated out<sup>8,16,17</sup>, which is further split into the Kubo cluster-cumulant functions<sup>18</sup>. The cluster-cumulant functions can be identified with particular components of the effective energy and are, finally, expanded into the generalized cumulant series<sup>18</sup>, this enabling us to obtain analytical expressions for the effective energy terms, including the very important multibody terms<sup>16,17</sup>. Owing to the application of this rigorous methodology, the number of interaction sites in UNRES and NARES-2P is only 2 per residue, while successful unrestrained simulations of the structure and dynamics of proteins and nucleic acids are successful with these models<sup>13,19</sup>.

UNRES has proved to be a successful tool for protein structure prediction, as assessed in the Community Wide Experiments on the Critical Assessment of Techniques for Protein Structure Prediction (CASP)<sup>20,21</sup>, and has been a useful tool for studying protein complexes<sup>22,23</sup>. NARES-2P<sup>13–15</sup> can treat both DNA and RNA systems<sup>14</sup>. The NARES-2P model was able to reproduce many biologically-relevant properties of double-helix B-DNA, such as duplex formation, breathing motions, melting temperatures<sup>14,15</sup>, and mechanical stability<sup>17</sup>.

Although the structure, dynamics, and thermodynamics of proteins and nucleic acids can be modeled reliably with UNRES and NARES-2P, until now, these two models did not cover systems containing both kinds of these biological macromolecules. Interactions between proteins and nucleic acids are essential for proper cell functioning<sup>24</sup>, DNA replication<sup>25</sup>, transcription activation<sup>26</sup> and regulation<sup>27</sup>, recombination process<sup>28,29</sup>, DNA repairing<sup>29,30</sup>, etc. Therefore, in this work, we combined the UNRES and NARES-2P models to handle protein – nucleic-acid systems. We used the coarse-grained potentials for the interactions between the protein and nucleic-acid components that we developed in our earlier work<sup>31</sup>. We integrated the UNRES and NARES-2P models and the protein – nucleic-acid interaction potentials into the newest UNRES package written in FORTRAN 90<sup>32</sup>. As for UNRES<sup>33,34</sup> and NARES-2P<sup>14</sup>, we implemented molecular dynamics (MD) as the main conformational-search engine<sup>33–35</sup>. After testing the implementation for energy and temperature conservation in microcanonical and canonical MD runs, respectively, we applied the newly developed model to study the initial stage of DNA-repair mechanism by the Ku70/Ku80 protein heterodimer<sup>30</sup> and found that the heterodimer is actively involved in DNA unwinding. We also applied the model to study conformational changes upon DNA - protein association with the example of multiple antibiotic resistance protein (MarA)<sup>36</sup>. We found that domain closing and major groove binding loop re-adjustment are the most important moves during DNA binding.

## METHODS

### Structure of the coarse-grained model of protein – nucleic-acid systems

In the model of protein – nucleic-acid systems developed in this work, the UNRES<sup>12,13,16,37–40</sup> representation is used for the protein part of a protein – nucleic acid-

system, while the NARES-2P<sup>13,14</sup> representation is used for the nucleic-acid part. As in UNRES and NARES-2P, the solvent is implicit in the model and is accounted for in the effective potentials of interactions between the coarse-grained sites. The effective energy function consists of the UNRES ( $U_{prot}^{UNRES}$ ) and the NARES-2P ( $U_{nucl}^{NARES-2P}$ ) energies of the protein and the nucleic-acid parts, respectively, and the effective energy of interactions between these parts ( $U_{prot-nucl}$ ), as expressed by eq. (1). It should be noted that both the protein and the nucleic-acid parts usually consist of multiple chains.

$$U = U_{prot}^{UNRES} + U_{nucl}^{NARES-2P} + U_{prot-nucl} \quad (1)$$

The UNRES and NARES-2P models and the pertinent effective energy functions are summarized in sections “UNRES model of polypeptide chains” and “NARES-2P model of the polynucleotide chains”, respectively, while the effective potentials of protein – nucleic-acid interactions are discussed in section “Effective potentials for the polypeptide-nucleic acid interactions”.

### UNRES model of polypeptide chains

In the UNRES model<sup>12,13,16,37–40</sup>, a polypeptide chain is represented by a sequence of  $\alpha$ -carbon ( $C^\alpha$ ) atoms, with united peptide groups (p) placed halfway between the two consecutive  $C^\alpha$  atoms, and united side chains (SC) attached to the  $C^\alpha$ s. Only the SC and p centers are interaction sites, while the  $C^\alpha$  atoms serve only to define the geometry of a polypeptide chain (Figure 1A).

The UNRES force field originates from the potential of mean force of a protein in aqueous environment, which has been expanded into a cluster-cumulant series to give an implementable effective energy function<sup>16,17</sup>. This energy function is given by eq. 2.

$$\begin{aligned}
U_{prot}^{UNRES} = & w_{SC} \sum_{i < j} U_{SC_i SC_j} + w_{SC} \sum_{p_i \neq j} U_{SC_i p_j} + w_{pp}^{VDW} \sum_{i < j-1} U_{p_i p_j}^{VDW} + w_{pp}^{el} f_2(T) \sum_{i < j-1} U_{p_i p_j}^{el} \\
& + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i) + w_{tor} f_3(T) \sum_i U_{tor}(\gamma_i, \gamma_{i+1}) \\
& + w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + w_{bond} \sum_i U_{bond}(d_i) \\
& + w_{corr}^{(3)} f_3(T) U_{corr}^{(3)} + w_{corr}^{(4)} f_4(T) U_{corr}^{(4)} + w_{turn}^{(3)} f_3(T) U_{turn}^{(3)} + w_{turn}^{(4)} f_4(T) U_{turn}^{(4)} \\
& + w_{SC-corr} f_2(T) \sum_{m=1}^3 \sum_i U_{SC-corr}(\tau_i^{(m)})
\end{aligned}
\tag{2}$$

where the  $U$ 's are energy terms,  $\theta_i$  is the backbone virtual-bond angle between the  $C_{i-1}^\alpha$ ,  $C_i^\alpha$  and  $C_{i+1}^\alpha$  atoms,  $\gamma_i$  is the backbone virtual-bond-dihedral angle, defined by the  $C_{i-1}^\alpha$ ,  $C_i^\alpha$ ,  $C_{i+1}^\alpha$ , and  $C_{i+1}^\alpha$  atoms,  $\alpha_i$  and  $\beta_i$  are the angles defining the location of the center of the united side chain of residue  $i$  (Figure 1A) with respect to the frame defined by the  $C_{i-1}^\alpha$ ,  $C_i^\alpha$  and  $C_{i+1}^\alpha$  atoms,  $d_i$  is the length of the  $i$ th virtual bond, which is either a  $C^\alpha \dots C^\alpha$  or a  $C^\alpha \dots SC$  virtual bond, and the angles  $\tau^{(1)} - \tau^{(3)}$  are the  $SC \dots C^\alpha \dots C^\alpha \dots C^\alpha$  ( $\tau^{(1)}$ ),  $C^\alpha \dots C^\alpha \dots C^\alpha \dots SC$  ( $\tau^{(2)}$ ), and  $SC \dots C^\alpha \dots C^\alpha \dots SC$  ( $\tau^{(3)}$ ), respectively. Each energy term is multiplied by the appropriate weight,  $w_x$ , and the terms corresponding to factors of order higher than 1 are additionally multiplied by the respective temperature factors which were introduced in our earlier work<sup>39</sup> and which reflect the dependence of the first generalized-cumulant term in those factors on temperature, as discussed in refs 39 and 41. The factors  $f_n$  are defined by eq. (3).

$$f_n(T) = \frac{\ln[\exp(1) + \exp(-1)]}{\ln\{\exp[(T/T_0)^n - 1] + \exp[-(T/T_0)^n - 1]\}} \tag{3}$$

where  $T_0=300$  K.

The terms  $U_{SC_i SC_j}$  represent the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains, which implicitly contain the contributions from the interactions of the side chains with the solvent. The terms  $U_{SC_i p_j}$  denote the excluded-volume potentials of the side-chain – peptide-group interactions. The peptide-group

interaction potentials are split into two parts: the Lennard-Jones terms  $\left(U_{p_i p_j}^{VDW}\right)$  and the average peptide-group-dipole interaction terms  $\left(U_{p_i p_j}^{el}\right)$ ; the second of these terms accounts for the tendency to form backbone hydrogen bonds between peptide groups  $p_i$  and  $p_j$ . The terms  $U_{tor}$ ,  $U_{torb}$ ,  $U_b$ ,  $U_{rob}$  and  $U_{bond}$  are the virtual-bond-dihedral angle torsional terms, virtual-bond dihedral angle double-torsional terms, virtual-bond angle bending terms, side-chain rotamer, and virtual-bond-deformation terms; these terms account for the local properties of the polypeptide chain. The terms  $U_{corr}^{(m)}$  represent correlation or multibody contributions from the coupling between backbone-local and backbone-electrostatic interactions, and the terms  $U_{turn}^{(m)}$  are correlation contributions that involve  $m$  consecutive peptide groups; they are, therefore, termed turn contributions. The multibody terms are indispensable for reproduction of regular  $\alpha$ -helical and  $\beta$ -sheet structures<sup>16,42,43</sup>. The  $U_{SC-corr}$  terms are side-chain backbone correlation potentials<sup>44,45</sup> introduced to the UNRES force field, which improved loop structures and secondary structure recognition of the UNRES force field.

The energy-term weights have been determined and other force-field parameters have been refined by calibrating the force field with a set of seven proteins that belong to different structural classes<sup>46</sup>.

### NARES-2P model of the polynucleotide chains

In the NARES-2P model<sup>13,14</sup>, a polynucleotide chain is represented by a sequence of deoxyribose (for DNA) or ribose (for RNA) ring centers (S) with united sugar – nucleic-acid-base (B) centers attached to them and united phosphate groups (P) located in the middle between the consecutive sugar-ring centers (Fig. 1B). As the C $\alpha$ s in UNRES, the sugar centers serve only to define the virtual-chain geometry, while the B and P centers are the interaction sites. The effective energy function of the NARES model<sup>14</sup> is expressed by eq. (4).

$$\begin{aligned}
 U_{nucl}^{NARES} = & w_{BB}^{GB} \sum_i \sum_{i < j} U_{B_i B_j}^{GB} + w_{BB}^{dip} f_2(T) \sum_i \sum_{i < j} U_{B_i B_j}^{dip} \\
 & + w_{PP} \sum_i \sum_{i < j} U_{P_i P_j} + w_{BP} \sum_i \sum_j U_{P_i B_j} + w_{bond} \sum_i U_{bond}(d_i) \\
 & + w_{bend} \sum_i U_{ang}(\theta_i) + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i) + w_{rot} \sum_i U_{rot}(\alpha_i, \beta_i)
 \end{aligned} \quad (4)$$

where  $U_{B_i B_j}^{GB}$  and  $U_{B_i B_j}^{dip}$  denote the Gay-Berne and mean field dipole-dipole interaction potentials of the sugar-base sites, respectively,  $U_{P_i P_j}$  and  $U_{P_i B_j}$  denote the phosphate-phosphate and phosphate – sugar-base interaction potentials, respectively,  $U_{bond}$ ,  $U_{ang}$ ,  $U_{tor}$  and  $U_{rot}$  denote virtual-bond-stretching, virtual-bond-angle-bending, torsional, and sugar-base-rotamer potentials, respectively, and  $f_2(T)$  is defined by eq. (3) with  $n = 2$ . Each term is

multiplied by the appropriate weight ( $w_x$ ); the weights have been determined by force-field calibration with small DNA molecules<sup>14,15</sup>.

This energy function was implemented in the new FORTRAN 90 UNRES package<sup>32</sup>.

### Effective potentials for the polypeptide-nucleic acid interactions

The energy function for DNA-nucleic acid interactions ( $U_{prot-nucl}$ ) is based on the potentials determined in our earlier work<sup>31</sup> and is expressed by eq. (5):

$$U_{prot-nucl} = w_{SC-B} \sum_i \sum_j U_{SC_i B_j} + w_{p-P} \sum_i \sum_j U_{p_i P_j} \quad (5) \\ + w_{p-B} \sum_i \sum_j U_{p_i B_j} + w_{SC-P} \sum_i \sum_j U_{SC_i P_j}$$

where  $U_{SC_i B_j}$ ,  $U_{p_i P_j}$ ,  $U_{p_i B_j}$  and  $U_{SC_i P_j}$  are protein side-chain – nucleic acid sugar-base, protein peptide-group – nucleic acid phosphate group, protein peptide-group – nucleic acid sugar-base, and protein side-chain – nucleic acid phosphate group effective interaction potentials, respectively, which are expressed by eqs. (S1-S53) of the Supporting Information of ref 31. Each term is multiplied by the appropriate weight ( $w_x$ ). Currently, all weights are set at 1. In order to determine the energy-term weights and fine-tune the other parameters of the force field calibration should be carried out<sup>15,46</sup> with a set of model small protein-DNA systems. This task requires a lot of effort and, therefore, we defer it to our further work. Nevertheless, as the two examples presented in the Results section suggest, even the present force field can give reliable results regarding the DNA-binding mode and functionally important motions of protein-DNA complexes.

Because we have found that charged side chains interact with the phosphate groups too strongly, the expressions for the electrostatic-interaction energies of the Asp-P, Glu-P, Lys-P, and Arg-P pairs have been modified in this work to introduce a Debye screening factor, as expressed by eq. (6). The other parameters of these potentials have been taken from our earlier work<sup>31</sup>.

$$E_{el} = 332 \frac{q_i q_j}{\epsilon_{in} R_{ij}} e^{-\kappa R_{ij}} \quad (6)$$

with

$$\kappa = \frac{1}{\sigma_i + \sigma_j} \quad (7)$$

where  $q_i = \pm 1$  is the charge of a side-chain headgroup,  $q_j = -1$  is the charge of the phosphate group,  $R_{ij}$  is the distance between the side-chain headgroup and the phosphate-group center,  $\epsilon_{in}$  is the effective dielectric constant, which was one of the fitted parameters in deriving the

potentials<sup>31</sup>,  $\kappa$  is the Debye screening constant, and  $\sigma_i$  is the thickness of the solvation shell of site  $i$  [see eq. (S27) of the Supporting Information of ref 31 and eq. (17) of ref 47]. Another modification was to reduce the distances of the polar head groups of the His and Asn residues from the side-chain centers to 1.74 Å and 1.01 Å, respectively because the distances obtained by PMF fitting were too long (4.62 Å and 6.04 Å, respectively; see Table S24 of the Supporting Information of ref 31).

The protein-DNA energy function described above was implemented in the new FORTRAN 90 package<sup>32</sup>.

### Tests of coarse-grained molecular dynamics

To verify the stability of MD runs with the new potentials for protein – nucleic-acid interactions, energy conservation in microcanonical MD simulations were checked first with the example of the MarA protein bound to DNA (PDB code: 1BL0)<sup>36</sup>. The system was energy-minimized before performing simulations, and 1,000,000 MD steps with 0.498 fs time step were subsequently run.

Temperature conservation with the new potentials was verified by running canonical simulations with the Langevin<sup>34</sup> and then Berendsen<sup>48</sup> thermostats implemented with UNRES<sup>34</sup>. The system was energy-minimized before performing simulations, and 1,000,000 MD steps with 4.98 fs time step were subsequently run. Both Langevin- and Berendsen-thermostat runs were performed at T=300 K. For the Berendsen thermostat<sup>48</sup>, the coupling parameter  $\tau = 48.9$  fs was used and water friction was scaled by the factor of 0.01 in Langevin simulations to speed up the calculations, as in our earlier work<sup>34</sup>.

### Dynamics of the Ku heterodimer bound to DNA

To study the initial stage of the DNA repair mechanism, MD simulations were carried out with the Ku heterodimer bound to damaged DNA (with mismatched base pairs). The starting structure was the crystal structure of the Ku70/Ku80 heterodimer bound to damaged DNA (PDB code: 1JEY)<sup>30</sup>. Because two protein loops are missing in the crystal structure, MODELLER<sup>49</sup> was used to add these missing sections of the structure. A series of five Langevin dynamics simulations at T = 300 K, 180,000 MD steps with 0.498 fs step length each, were carried out. These simulations were run with a small time step in order to capture all important events that take place during DNA repair by the Ku protein. It should be kept in mind that the time scale of coarse-grained simulations is at least 1,000 times wider than that of all-atom simulations<sup>34,50</sup> and, consequently, processes occur much faster at the coarse-grained level. This time-scale extension results from the high reduction of the number of explicitly treated degrees of freedom (only 6 virtual-bond-vector coordinates per residue<sup>34</sup>) and smoothing the free energy landscape. Because of this time-scale distortion, the simulation settings correspond to approximately 90 ns of an all-atom simulation. Snapshots were collected every 1000 steps (a total of 900 snapshots from all trajectories). To select a representative structure, a set composed of the last structures from all trajectories was created first. The structure with the lowest maximum root-mean-square deviation (RMSD) from all other structures was then selected as a representative structure. This representative structure was subsequently superposed on the experimental structure and the



deviations of the consecutive residues of the protein and DNA parts from their experimental counterparts were calculated and plotted in residue index. Subsequently, the structures from all 900 snapshots were subjected to Principal Component Analysis (PCA)<sup>51–53</sup> to determine the most important motions that characterize the initial stages of the DNA-repair mechanism.

### Dynamics of multiple antibiotic resistance protein (MarA) during DNA binding

To study the conformational changes of the MarA protein during the DNA binding, the crystal structure of the MarA – DNA complex was energy-minimized first. The protein and DNA chains were subsequently separated from each other by shifting to the distance of 5 Å and by 8 Å to produce the starting structure. Two series of 16 Langevin dynamics trajectories and 200,000 steps per trajectory each were run. The simulations of the first series were started with the initial configuration in which the DNA molecule was separated by 5 Å from the protein; the time step in these series of simulations was set at 0.489 fs, this corresponding to a total of 100 ns all-atom-simulation time (given an approximately 1000-fold time-scale extension<sup>34</sup>). In the second series of simulations, DNA was initially at a larger, 8 Å, distance from the protein and the simulations were carried out with the time step of 4.89 fs, which corresponded to approximately 100 ns simulation time per trajectory. Each simulation was run in the NVT mode at the temperature of  $T = 300$  K. The time step was reduced to 0.498 fs in the first simulation series, because the binding of DNA to MarA occurred very fast in the coarse-grained model at such a small initial separation of the components, and setting a larger time step lead to omission of important events that occur upon binding. Snapshots were collected every 1,000 steps (a total of 3,200 snapshots from all trajectories). For each series of simulations, a representative structure was selected as described in section “Dynamics of the Ku heterodimer bound to DNA”.

In order to get better insight into the DNA-protein binding pattern, contact maps were calculated for both series of simulations. For each series of simulations, the distances between the  $C^\alpha$  atoms of the protein and the  $C5'$  atoms of the DNA were calculated. For a given  $C^\alpha \cdots C5'$  pair, a score of 1 was assigned if the respective distance was 8 Å or less, 0, if it was 10 Å or greater, while the score decreased linearly from 1 to 0 for distances between 8 Å and 10 Å. Then the scores were averaged over 10 last snapshots (a total of 160 snapshot for each simulation series). The last structures from all trajectories were compared with each other. The structure with the lowest maximum root-mean-square deviation (RMSD) from all the others selected structures was selected as a representative structure. The results of the simulation in which the initial separation of the protein and DNA molecules was 5 Å were used in the analysis and in the principal component analysis described in the next section. Residue-wise deviations of the protein and DNA parts of the representative structure from the experimental structure after optimal superposition were calculated and the corresponding residue-wise plots were constructed. Subsequently, all 3,200 snapshot structures were subjected to PCA<sup>51–53</sup> to determine the most important motions that occur during DNA-protein association.

## Principal component analysis

PCA is an implementation of the Normal Mode Analysis (NMA) and is an efficient method to identify all collective motions of a molecular system<sup>51–53</sup>. A key concept of NMA is that the displacements of atom/site coordinates of the molecular system from those of the average structure are expressed as a linear combination of normal mode coordinates. The normal modes are obtained as eigenvectors of the variance-covariance matrix of coordinates  $\mathbf{C}$  defined by eq. (8),

$$C_{ij} = \frac{1}{N-1} \sum_{k=1}^N (q_{ik} - \langle q_i \rangle)(q_{jk} - \langle q_j \rangle), \quad i = 1, 2, \dots, 3n, \quad j = 1, 2, \dots, 3n \quad (8)$$

where  $q_{ik}$  is the mass-scaled  $x$ ,  $y$  or  $z$  coordinate of the respective atom/site of the  $k$ th snapshot defined by eq. (9),  $\langle q_i \rangle$  is the respective average coordinate (the coordinates are averaged after all structures are superposed on a selected structure of the trajectory),  $N$  is the number of snapshots and  $n$  is the total number of atoms/sites in the system.

$$\begin{aligned} q_{3(i-1)+1,k} &= m_i^{\frac{1}{2}} x_i \\ q_{3(i-1)+2,k} &= m_i^{\frac{1}{2}} y_i \\ q_{3i,k} &= m_i^{\frac{1}{2}} z_i \end{aligned} \quad (9)$$

where  $m_i$  is the mass of the  $i$ th atom/site. The eigenvectors of the matrix  $\mathbf{C}$ , denoted as  $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_{3n}$ , respectively, and termed the *principal components*, are arranged according to the descending order of the corresponding eigenvalues,  $\chi_i$ ,  $i = 1, 2, \dots, 3n$ , so that  $\chi_1 > \chi_2 > \dots > \chi_{3n}$ . The top-ranked PC modes describe global motions of the system, while the low ranked PC modes describe the local motions. The eigenvalues normalized to the sum of 1 are termed *loadings*. The ProDy interface of the Normal mode visualization and comparative analysis (NMWiz) plug-in of the Visual Molecular Dynamics (VMD) software was used to run PCA<sup>54,55</sup>.

## RESULTS

### Tests of MD behavior in microcanonical and canonical runs

Variation of the total energy with MD simulation time for the MarA-DNA system is shown in Figure 2A. As can be seen from the Figure, the total energy fluctuates only about the average value, the variations amounting to 0.008 kcal/mol, this proving symplectic behavior of the simulation.

The plots of temperature distribution for the Langevin and Berendsen run are shown in Figure 2B. It can be seen that, for the Langevin run, the distribution resulting from the simulations matches the theoretical distribution for the respective number of degrees of freedom, while the Berendsen thermostat produces too narrow a distribution, as found in

earlier work<sup>56,57</sup>. Nevertheless, the mean temperature is very close to the bath temperature for both simulations.

### Dynamics of Ku heterodimer bound to DNA

The structure of the Ku heterodimer bound to DNA after simulations is compared with the experimental structure in Figure 3. As can be seen from the Figure, the most significant changes occurred at the wrongly paired fragment of damaged DNA. The pairing was broken and one of the DNA chains moved towards the protein. The conformational changes are also visualized in Figure 4, where the residue-wise deviations from the experimental structure are shown. As can be seen from Figure 4B, large deviations are observed for the shorter DNA chain at the C3' end and at the C5' end, and in the loop region of the longer chain. The RMSD of the representative final structure from the experimental structure is 5.19 Å whereas, for the DNA part, the RMSD is 6.71 Å after superposing the DNA fragments of the system. For the Ku protein, the largest fluctuations are observed in the  $\beta$ -sheet of Ku70 that “grabs” the DNA strands and in the C-terminal fragment of Ku70.

The loadings (normalized eigenvalues) corresponding to the protein, the nucleic-acid and the protein – nucleic-acid-complex are plotted in Figure S1 of the Supporting Information, while the principal components, together with the plots of the fluctuations of the  $C^\alpha$  atoms along the principal components, are shown in Figure 5 and Figures S2 and S3 of the Supporting Information. The results of PCA enable us to conclude that the most important motions of the protein consist of (i) reorganization of the loops (residues 289–295 and 304–312 of Ku70, and residues 280–287 and 298–304 of Ku80) in the vicinity of the the  $\beta$ -sheet (Figure 5D), which is composed of one strand from Ku70 (residues 289–297) and one from Ku80 (residues 296–303), (ii) “grabbing” the damaged DNA by the  $\beta$ -sheet (Figure 5A), this motion corresponding to the PC mode with the largest loading, (iii) opening the wrongly-paired DNA by the C-terminal fragment of Ku70 and the N-terminal part of Ku80 (Figure 5C), and (iv) global motion of the protein domains (Figure 5B and E). The most important DNA motions (Supplementary Figures S3) are those corresponding to the wrongly formed loop and wrongly paired DNA-fragment. When this system is analyzed as a whole (Figure 5) it is found that the most important motions are well coordinated and both chains act in harmony in the DNA-repair process.

To confirm the motion pattern detected by coarse-grained simulations, we have also carried all-atom MD simulations of the Ku/DNA system with the AMBER force field<sup>58,59</sup>. The simulations were carried out in the NPT regime, at a temperature of  $T = 298$  K and pressure of  $p = 1$  atm, in a periodic-water box with initial dimensions  $100 \times 105 \times 130$  Å containing 40,000 TIP3P water molecules and 12  $\text{Na}^+$  counter-ions to zero out the net charge. Four independent trajectories, 5 ns each were run. All data were used in principal component analysis. The resulting principal components, together with the plots of the fluctuations of the  $C^\alpha$  atoms along the principal components, are visualized in Figure 6. It can be seen that the components corresponding to grabbing (panels A and B) and opening (panel C) the damaged DNA section are similar to their counterparts from coarse-grained simulations (Figure 5). However, the motion of the loops at residues 289–295 and 304–312 of Ku70, and residues 280–287 and 298–304 of Ku80 are less pronounced probably because the motion of

the crowded section of the system is facilitated in coarse-grained representation in which the effective energy surface is smoother. Additionally, the  $\alpha$ -helix at residues 140–152 moves, which motion is not seen in the coarse-grained dynamics.

Our results suggest that the Ku complex is actively involved in the unwinding of wrongly paired DNA fragments. These results are consistent with the experimentally determined active role of this complex<sup>60</sup>.

### Dynamics of multiple antibiotic resistance protein (MarA) during the DNA binding

As mentioned in Methods, two series of simulations, one started from initial protein-DNA separation of 5 Å and one from the separation of 8 Å were carried out. The DNA-protein complex was formed in all 16 trajectories for both series of simulations. The crystal and the representative structure from the end of the simulations starting from 5 Å separation of the MarA-DNA complex are shown in panels A and B of Figure 7, respectively. It can be seen from the Figure that the protein is bound to DNA in a manner similar to that of the crystal structure<sup>36</sup>. The RMSD between the representative structure and the crystal structure is 6.68 Å. The RMSD of the protein part is 4.78 Å, while that of the nucleic-acid part is 7.57 Å. It should be noted that, because of a larger dimension of a nucleotide than that of an amino-acid residue, the RMSD of 7.57 Å for the nucleic-acid part means that the structure superposes quite well on its experimental counterpart. On the other hand, the components are packed too tightly in the simulated complex (Figure 7B).

In order to get better insight into the binding mode of DNA to MarA, we analyzed the contacts between its component in the experimental structure<sup>36</sup> and average contacts in the simulated structure. The respective contact maps are shown in Figure 8A–C. From the analysis of the contacts in the experimental structure (Figure 8A), it can be seen that residues 35–55 and 80–100, which correspond to the loop regions of MarA, are the most important for DNA binding; these residues bind to the major groove of the DNA molecule. The averaged contact maps obtained from simulations indicate the same general binding pattern (Figure 8B and C); however, the contacts are more diffuse and additional contacts involving the residues engaged in the native contacts and the neighboring residues are formed. The contacts corresponding to the simulations in which the MarA and DNA molecules were initially separated by 8 Å are even more diffuse and the contact scores are lower; however, the binding follows the native mode (Figure 8C). These results suggest that the binding process is dynamic.

The residue-wise deviations of the representative final simulated structure from the experimental structure (Figure 9) indicate that the greatest deviations occur in the C5' fragment of both DNA chains and slightly smaller deviations at the C3' end. The middle part of DNA is stable. For the protein, the most significant changes occur at the major groove binding loop (residues 20–45, 29–54 according to residue numbering in the PDB file).

The loadings of the normal modes for the protein, nucleic-acid and the complex are plotted in Figure S4 of the Supporting Information, while the principal components are visualized in Figure 10 and Figures S5 and S6 of the Supporting Information. The principal components indicate that the most important motions during protein-DNA association are (i) the motion

of DNA towards the protein (Figure 10A), (ii) protein domain closing (Figure 10B) and (iii) the motion that corresponds to the binding of parts of the protein to the major groove (Figure 10C and D). These motions are well separated probably because DNA was separated from the protein in the starting structure.

## CONCLUSIONS

In this work, we developed a coarse-grained model of protein-DNA systems, which combines UNRES and NARES-2P, as well as the recently determined coarse-grained potentials for protein – nucleic-acid interactions<sup>31</sup>. The model was integrated into the UNRES software package for coarse-grained molecular dynamics<sup>32</sup>. The enhanced software run on model protein-DNA systems passed the energy-conservation tests in microcanonical simulations and also produced correct temperature distribution in test Langevin simulations and correct mean temperature in the Berendsen-thermostat simulations.

As the first applications of the coarse-grained model of proteins and nucleic acid systems developed in this work, we simulated the initial stage of the repair of damaged DNA by investigating the dynamics of the complex of the Ku protein heterodimer with damaged DNA and DNA binding to the MarA protein. We found that the initial stage of the repairing of damaged DNA involves breaking the wrong base pairing because of interactions with the Ku heterodimer. Additionally, we found that the DNA bulge also undergoes major structure reorganization. We also found that, in the simulated binding of DNA to MarA, the motions corresponding to binding the protein to the DNA are well separated from DNA association motions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

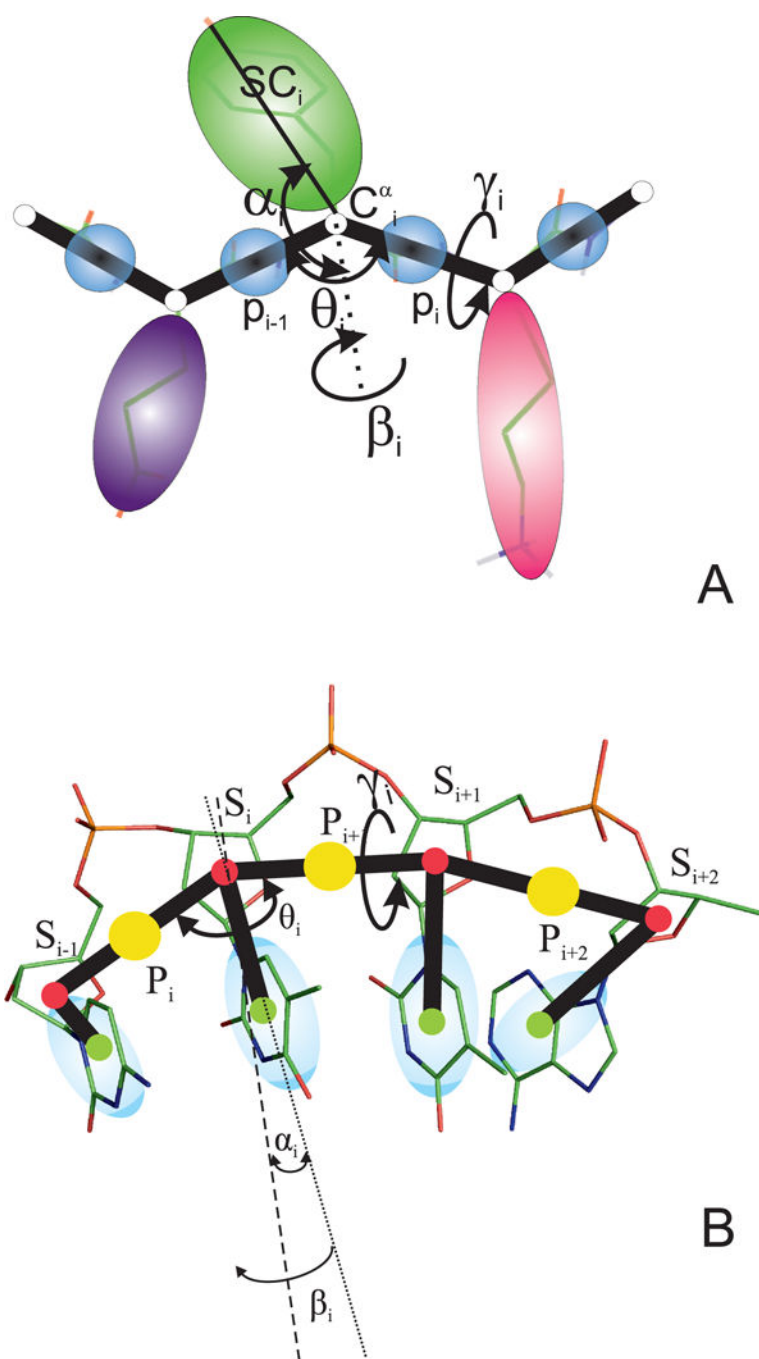
This work was supported by the National Science Center (Poland) Sonata UMO-2015/17/D/ST4/00509 (to AKS) and UMO-2012/06/A/ST4/00376 (to AL) and by NIH grant GM-14312 (to HAS, YY, YH). Computational resources were also provided by (a) the supercomputer resources at the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, and (b) computational resources at Interdisciplinary Center for Mathematical and Computer Modeling in Warsaw (ICM), grant GA71–21 (c) Polish Grid Infrastructure (PL-GRID) (d) our 682-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk (e) the 624-processor Beowulf cluster at the Baker Laboratory of Chemistry, Cornell University, U.S.A.

## References

1. Pande VS, Baker I, Chapman J, Elmer S, Kaliq S, Larson SM, Rhee YM, Shirts MR, Snow CD, Sorin EJ, et al., *Biopolymers* 68, 91 (2003). [PubMed: 12579582]
2. Hess B, Kutzner C, van der Spoel D, and Lindahl E, *J. Chem. Theory Comput* 4, 435 (2008). [PubMed: 26620784]
3. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, and Schulten K, *J. Comput. Chem* 26, 1781 (2005). [PubMed: 16222654]
4. Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD, et al., *ACM/IEEE SC 2006 Conference (SC.06)* pp. 43–43 (2006).
5. Friedrichs MS, Eastman P, Vaidyanathan V, Houston M, Legrand S, Beberg AL, Ensign DL, Bruns CM, and Pande VS, *J. Comput. Chem* 30, 864 (2009). [PubMed: 19191337]

6. Lindorff-Larsen K, Piana S, Dror RO, and Shaw DE, *Science* 334, 517 (2011). [PubMed: 22034434]
7. Dror RO, Dirks RM, Grossman JP, Xu H, and Shaw DE, *Annu. Rev. Biophys* 41, 429 (2012). [PubMed: 22577825]
8. Ayton GS, Noid WG, and Voth GA, *Curr. Opin. Struct. Biol* 17, 192 (2007).
9. Czaplewski C, Liwo A, Makowski M, Ołdziej S, and Scheraga HA, in *Multiscale Approaches to Protein Modeling*, edited by Koliński A (Springer, 2010), chap. 3, pp. 35–83.
10. Marrink SJ and Tieleman DP, *Chem. Soc. Rev* 42, 6801 (2013). [PubMed: 23708257]
11. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid A, and Koliński A, *Chem. Rev* 116, 7898 (2016). [PubMed: 27333362]
12. Liwo A, Czaplewski C, Ołdziej S, Rojas AV, Kamińskiewicz R, Makowski M, Murarka RK, and Scheraga HA, in *Coarse-Graining of Condensed Phase and Biomolecular Systems*, edited by Voth G (CRC Press, 2008), chap. 8, pp. 1391–1411.
13. Liwo A, Baranowski M, Czaplewski C, Gołaś E, He Y, Jagieła D, Krupa P, Maciejczyk M, Makowski M, Mozolewska MA, et al., *J. Mol. Model* 20, 2306 (2014). [PubMed: 25024008]
14. He Y, Maciejczyk M, Ołdziej S, Scheraga HA, and Liwo A, *Phys. Rev. Lett* 110, 098101 (2013). [PubMed: 23496746]
15. He Y, Liwo A, and Scheraga HA, *J. Chem. Phys* 143, 243111 (2015). [PubMed: 26723596]
16. Liwo A, Czaplewski C, Pillardy J, and Scheraga HA, *J. Chem. Phys* 115, 2323 (2001).
17. Sieradzan AK, Krupa P, and Wales DJ, *J. Phys. Chem. B* 121, 2207 (2017). [PubMed: 28194966]
18. Kubo R, *J. Phys. Soc. Japan* 17, 1100 (1962).
19. Baranowski M, Czaplewski C, Gołaś EI, He Y, Jagieła D, Krupa P, Liwo A, Maisuradze GG, Makowski M, Mozolewska MA, et al., in *Coarse-Grained Modeling of Biomolecules*, edited by Papoian GA (CRC Press, 2017), chap. 3, pp. 67–120.
20. He Y, Mozolewska MA, Krupa P, Sieradzan AK, Wirecki TK, Liwo A, Kachlishvili K, Rackovsky S, Jagieła D, Lusarz R, et al., *Proc. Nat. Acad. Sci. U.S.A* 110, 14936 (2013).
21. Krupa P, Mozolewska MA, Wiñiewska M, Yin Y, He Y, Sieradzan AK, Ganzynkiewicz R, Lipska AG, Karczyńska A, Lusarz M, et al., *Bioinformatics* 32, 3270 (2016). [PubMed: 27378298]
22. Sieradzan AK, Liwo A, and Hansmann UHE, *J. Chem. Theory Comput* 8, 3416 (2012). [PubMed: 24039552]
23. Mozolewska M, Krupa P, Scheraga H, and Liwo A, *Proteins: Struct., Funct., Bioinf* 83, 1414 (2015).
24. Roos WP and Kaina B, *Trends Mol. Med* 12, 440 (2006). [PubMed: 16899408]
25. Fumasoni M, Zwicky K, Vanoli F, Lopes M, and Branzei D, *Molecular Cell* 57, 812 (2015). [PubMed: 25661486]
26. Baeuerle PA, *Biochim. Biophys. Acta* 1072, 63 (1991). [PubMed: 2018779]
27. Jones PL, Veenstra GCJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, and Wolffe AP, *Nature Gen* 19, 187 (1998).
28. Longhese MP, Plevani P, and Lucchini G, *Mol. Cell. Biol* 14, 7884 (1994). [PubMed: 7969128]
29. Wold MS, *Annu. Rev. Biochem* 66, 61 (1997). [PubMed: 9242902]
30. Walker JR, Corpina RA, and Goldberg J, *Nature* 412, 607 (2001). [PubMed: 11493912]
31. Yin Y, Sieradzan AK, Liwo A, He Y, and Scheraga HA, *J. Chem. Theory Comput* 11, 1792 (2015). [PubMed: 26052263]
32. Lubecka E and Liwo A, *TASK Quart* 20, 399 (2016).
33. Khalili M, Liwo A, Rakowski F, Grochowski P, and Scheraga HA, *J. Phys. Chem. B* 109, 13785 (2005). [PubMed: 16852727]
34. Khalili M, Liwo A, Jagielska A, and Scheraga HA, *J. Phys. Chem. B* 109, 13798 (2005). [PubMed: 16852728]
35. Rakowski F, Grochowski P, Lesyng B, Liwo A, and Scheraga HA, *J. Chem. Phys* 125, 204107 (2006). [PubMed: 17144690]
36. Rhee S, Martin RG, Rosner JL, and Davies DR, *Proc. Natl. Acad. Sci. U. S. A* 95, 10413 (1998). [PubMed: 9724717]

37. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, and Scheraga HA, *Protein Sci* 2, 1715 (1993). [PubMed: 8251944]
38. Liwo A, Ołdziej S, Czaplewski C, Koz-lowska U, and Scheraga HA, *J. Phys. Chem. B* 108, 9421 (2004).
39. Liwo A, Khalili M, Czaplewski C, Kalinowski S, Ołdziej S, Wachucik K, and Scheraga H, *J. Phys. Chem. B* 111, 260 (2007). [PubMed: 17201450]
40. Liwo A, He Y, and Scheraga HA, *Phys. Chem. Chem. Phys* 13, 16890 (2011). [PubMed: 21643583]
41. Shen H, Liwo A, and Scheraga HA, *J. Phys. Chem. B* 113, 8738 (2009). [PubMed: 19480420]
42. Kolinski A and Skolnick J, *J. Chem. Phys* 97, 9412 (1992).
43. Liwo A, Kamińskie R, Czaplewski C, Groth M, Ołdziej S, Wawak RJ, Rackovsky S, Pincus MR, and Scheraga HA, *J. Comput. Chem* 19, 259 (1998).
44. Krupa P, Sieradzan AK, Rackovsky S, Baranowski M, Ołdziej S, Scheraga HA, Liwo A, and Czaplewski C, *J. Chem. Theory Comput* 9, 4620 (2013).
45. Sieradzan AK, Krupa P, Scheraga HA, Liwo A, and Czaplewski C, *J. Chem. Theory Comput* 11, 817 (2015). [PubMed: 25691834]
46. Krupa P, Hałabis A, Mudzińska W, Ołdziej S, Scheraga HA, and Liwo A, *J. Chem. Inf. Model* 57, 2364 (2017). [PubMed: 28809487]
47. Makowski M, Liwo A, and Scheraga HA, *J. Phys. Chem. B* 115, 6130 (2011). [PubMed: 21500791]
48. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, and Haak JR, *J. Chem. Phys* 81, 3684 (1984).
49. Webb B and Sali A, *Methods Mol. Biol* 1137, 1 (2014). [PubMed: 24573470]
50. Sieradzan AK, Scheraga HA, and Liwo A, *J. Chem. Theor. Comput* 8, 1334 (2012).
51. Cui Q and Bahar I, *Normal mode analysis: theory and applications to biological and chemical systems* (CRC press, 2005).
52. Maisuradze GG and Leitner DM, *Proteins: Struct. Func. Bioinfo* pp. 569–578 (2007).
53. Yang L-W, Eyal E, Bahar I, and Kitao A, *Bioinformatics* 25, 606 (2009). [PubMed: 19147661]
54. Humphrey W, Dalke A, and Schulten K, *J. Mol. Graph* 14, 33 (1996). [PubMed: 8744570]
55. Bakan A, Meireles LM, and Bahar I, *Bioinformatics* 27, 1575 (2011). [PubMed: 21471012]
56. D'Alessandro M, Tenenbaum A, and Amadei A, *J. Phys. Chem. B* 106, 5050 (2002).
57. Sieradzan AK, *J. Comput. Chem* 36, 940 (2015). [PubMed: 25753584]
58. Pearlman D, Case D, Caldwell J, Ross W, Cheatham T III, DeBolt S, Ferguson D, Seibel G, and Kollman P, *Comp. Phys. Commun* 91, 1 (1995).
59. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, et al., *Amber 12* (2012), University of California: San Francisco.
60. Ramsden DA and Gellert M, *EMBO J* 17, 609 (1998). [PubMed: 9430651]

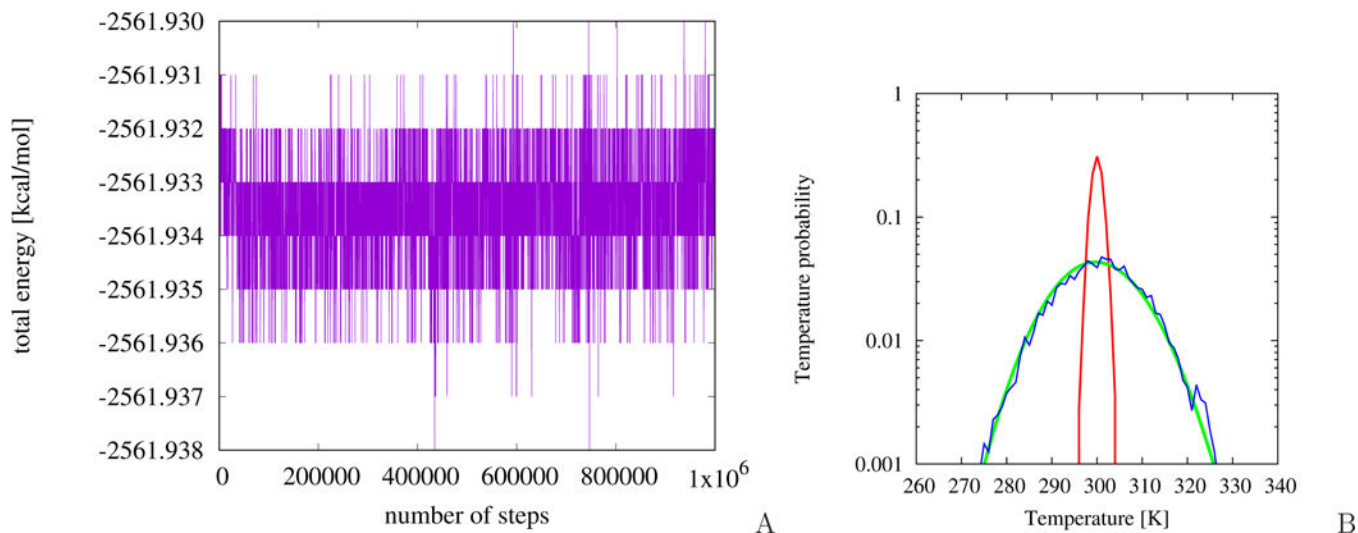


**Figure 1.**

(A) UNRES model of polypeptide chains. The interaction sites are united peptide groups (p) represented by light-blue spheres, which are located halfway between the consecutive  $\alpha$ -carbon atoms (shown as small white spheres), and united side chains attached to the  $\alpha$ -carbon atoms (SC), represented by spheroids with different colors and dimensions. Backbone geometry of the simplified polypeptide chain is defined by the  $C^\alpha \dots C^\alpha \dots C^\alpha$  virtual-bond angles  $\theta$  ( $\theta_i$  has the vertex at  $C^\alpha$ ) and the  $C^\alpha \dots C^\alpha \dots C^\alpha \dots C^\alpha$  virtual-bond-dihedral angles  $\gamma$  ( $\gamma_i$  has the axis passing through  $C_i^\alpha$  and  $C_{i+1}^\alpha$ ). The local geometry

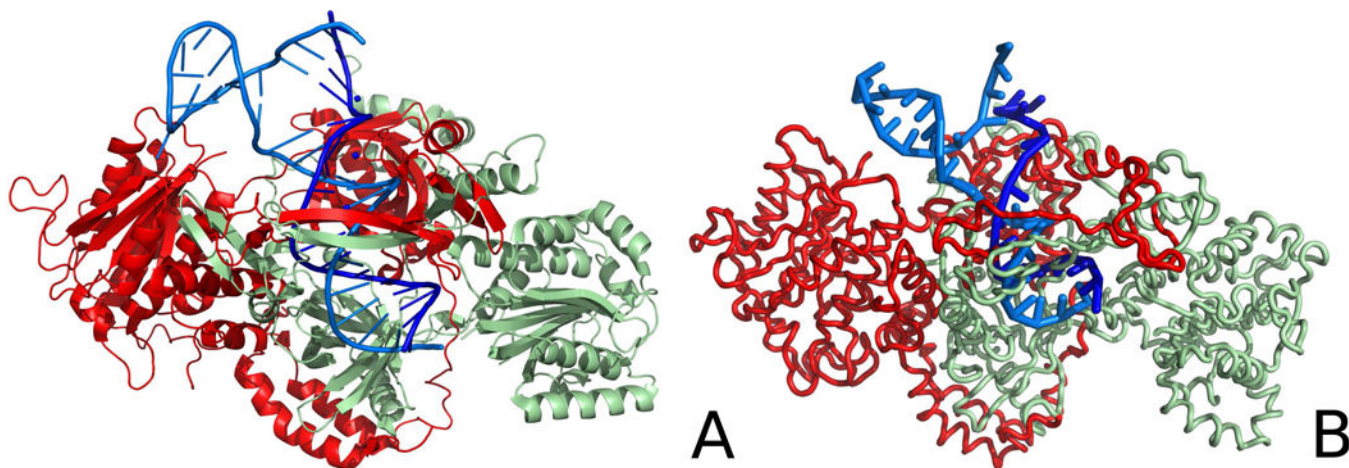


of the  $i$ th side-chain center is defined by the zenith angle  $\alpha_i$  (the angle between the bisector of the respective angle  $\theta_i$  and the  $C_i^\alpha \cdots SC_i$  vector) and the azimuthal angle  $\beta_i$  (the angle of counter-clockwise rotation of the  $C_i^\alpha \cdots SC_i$  vector about the bisector from the  $C_{i-1}^\alpha \cdots C_i^\alpha \cdots C_{i+1}^\alpha$  plane, starting from  $C_{i-1}^\alpha$ ). For illustration, the bonds of the all-atom chains, except for those to the hydrogen atoms connected with the carbon atoms, are superposed on the coarse-grained picture. (B) NARES-2P model of polynucleotide chains. The interaction sites are united phosphate groups (P), represented by yellow circles, which are located halfway between the centers of the consecutive sugar rings (S; shown as small red spheres), and united sugar-base groups (B), represented by light-blue spheroids. The united sugar centers serve only to define chain-backbone geometry. The meaning of angles,  $\theta$ ,  $\gamma$ ,  $\alpha$ , and  $\beta$  is as in panel A.

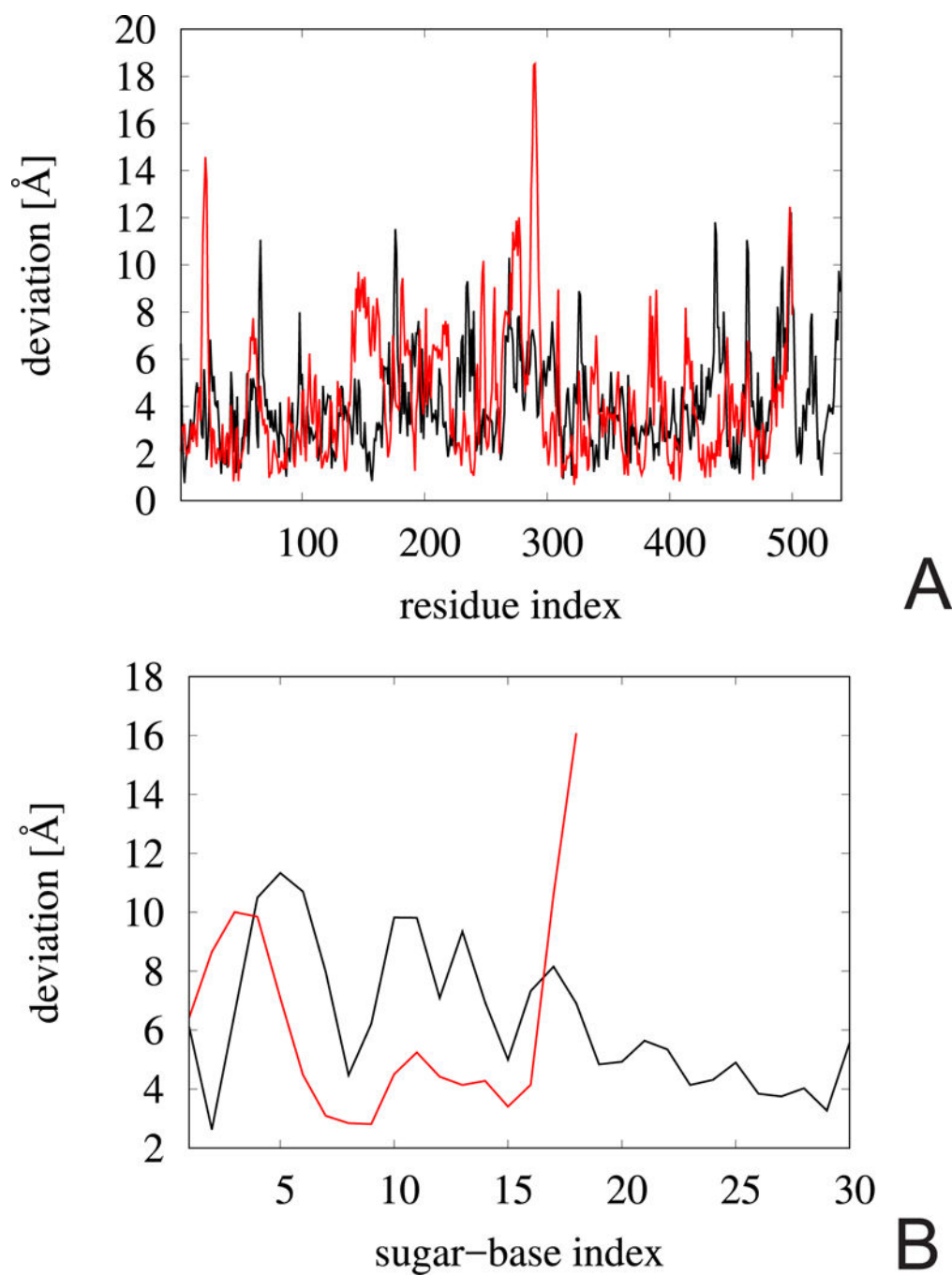


**Figure 2.**

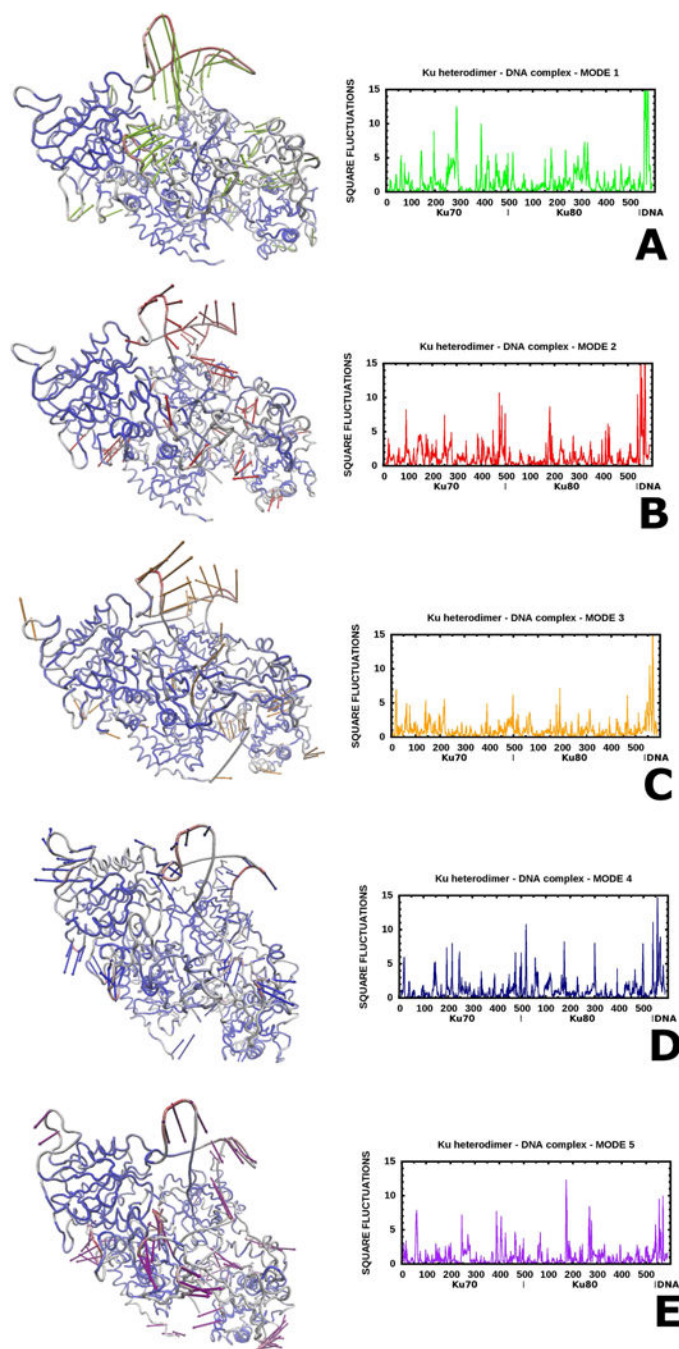
(A) Plot of the total energy changes as a function of time. Simulations were performed in micro-canonical mode for the MarA protein bound to DNA. (B) Temperature probability distribution plot (shown in logarithmic scale) for the Berendsen thermostat (red) and Langevin thermostat (blue) compared with theoretical distribution (green) given by the equation  $P(T) = AT^{\frac{g-2}{2}} \exp\left(-\frac{gT}{2T_b}\right)$ , where  $g$  is the number of the degrees of freedom,  $T_b$  is the thermal-bath temperature, and  $A$  is the normalization constant.



**Figure 3.** Comparison of the crystal structure of the Ku heterodimer bound to DNA (A) with the representative structure of this system obtained after coarse-grained MD simulation (B). The Ku70 (chain A) and Ku80 (chain B) monomers are colored pale-green and red, respectively, DNA chains C and D are colored dark blue and light blue, respectively. The chains are labeled as in the 1JEY experimental structure of the Ku protein.



**Figure 4.** Deviations of the positions of the  $C^\alpha$  and  $C5'$  atoms of the representative structure obtained in coarse-grained MD simulations of the DNA-Ku heterodimer from their positions in the experimental structure of the Ku protein (PDB code: 1JEY) as functions of residue index (A) and the respective deviations of the DNA bound to the Ku heterodimer (B). Red color: the Ku70 monomer (chain A) and DNA chain C, black color: the Ku80 monomer (chain B) and DNA chain D.



**Figure 5.** Illustration of the first five principal components (panel A-E) obtained by doing the PCA of the snapshots of the Ku heterodimer bound to DNA collected from the coarsegrained MD simulations. The principal components are represented by arrows colored green (A), red (B), orange (C), blue (D), and purple (E), respectively. The structure displayed in a given panel has been obtained by adding the respective eigenvector coordinates to the site coordinates; the deviations from the average structure are marked by coloring the respective parts of the system from blue (low deviations) to deep red (high deviations). On the right side of each

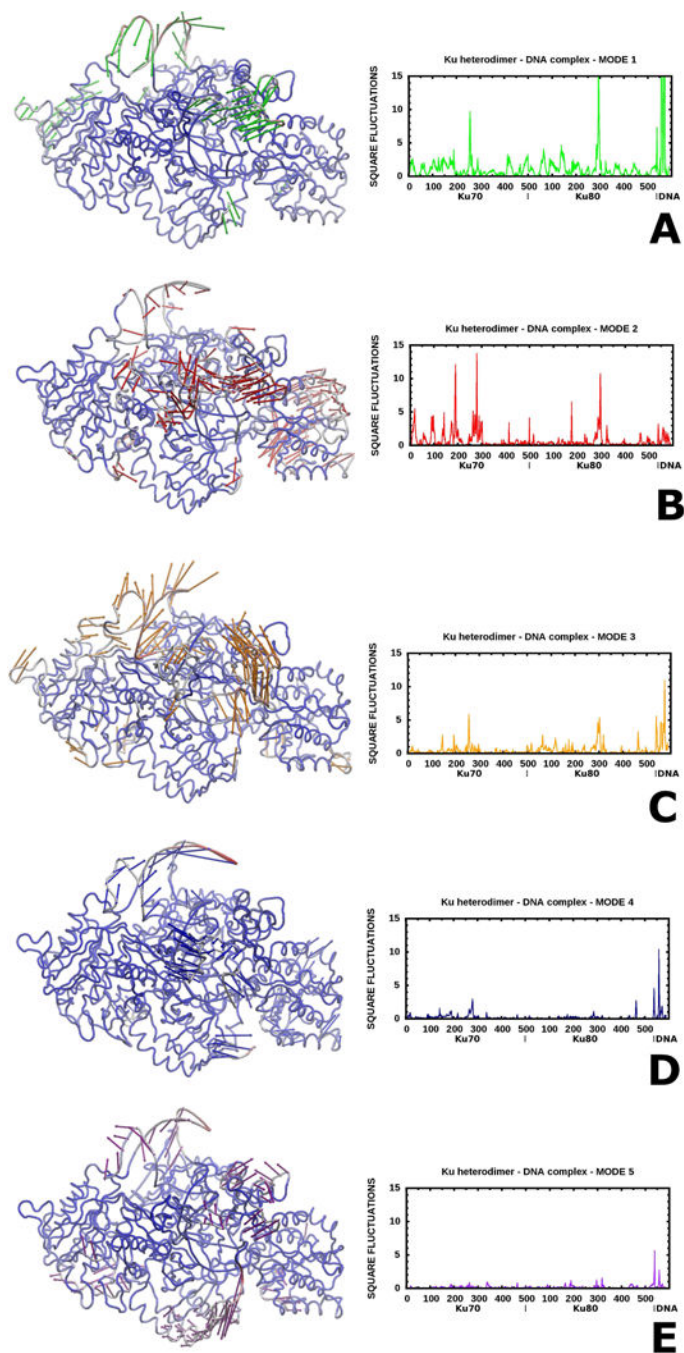
panels, the the squares of C <sup>$\alpha$</sup> -atom the fluctuations corresponding to a given component are plotted.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 6.** Illustration of the first two principal components (panels A-E) obtained by doing the PCA of the snapshots of the Ku heterodimer bound to DNA collected from the all-atom MD simulations. The principal components are represented by arrows colored green (A), red (B), orange (C), blue (D), and purple (E), respectively. The structure displayed in a given panel has been obtained by adding the respective eigenvector coordinates to the site coordinates; the deviations from the average structure are marked by coloring the respective parts of the system from blue (low deviations) to deep red (high deviations). On the right side of each

panels, the the squares of  $C^{\alpha}$ -atom the fluctuations corresponding to a given component are plotted.

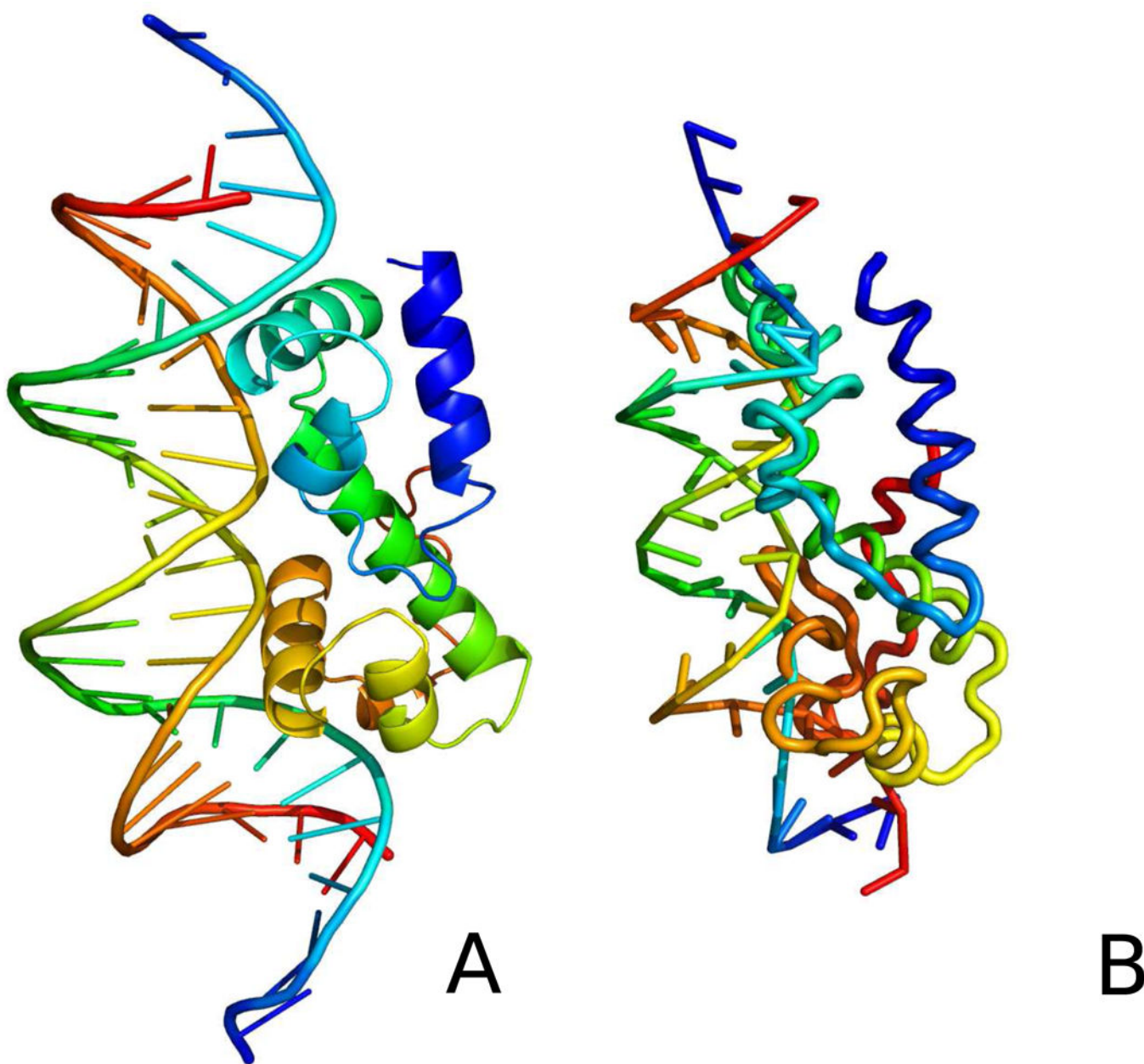
Author Manuscript

Author Manuscript

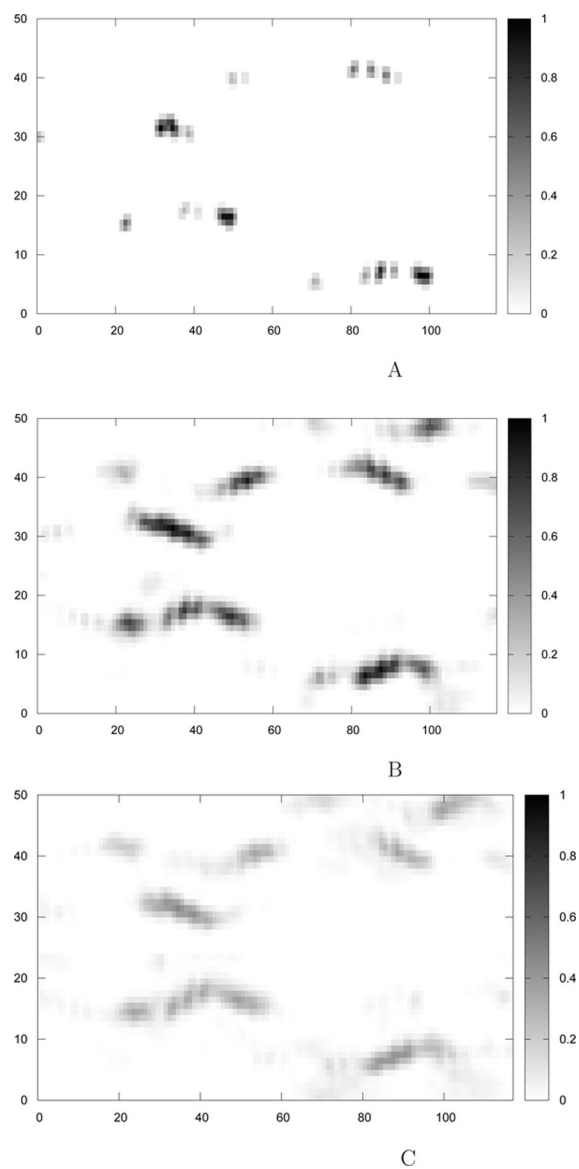
Author Manuscript

Author Manuscript

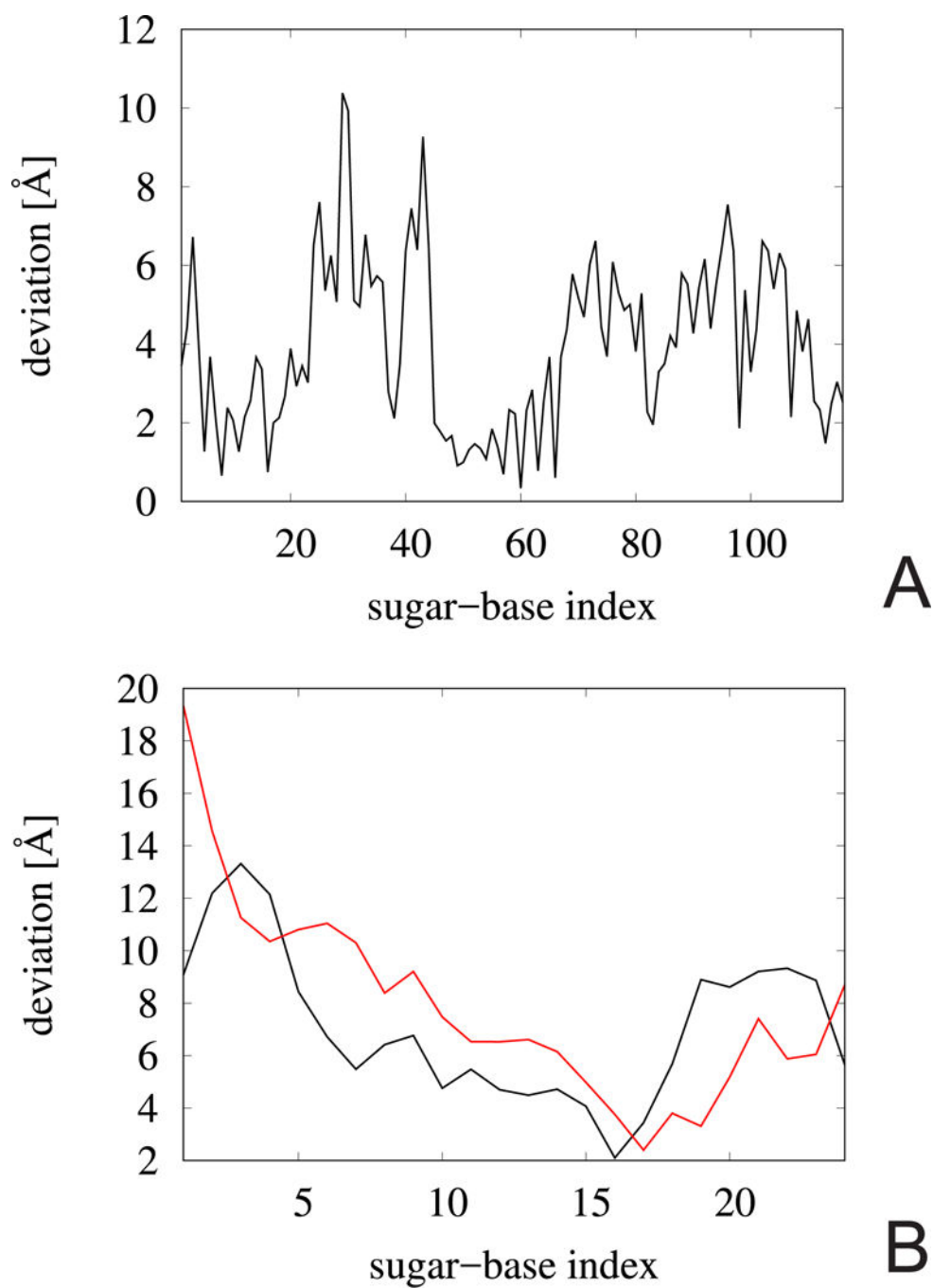




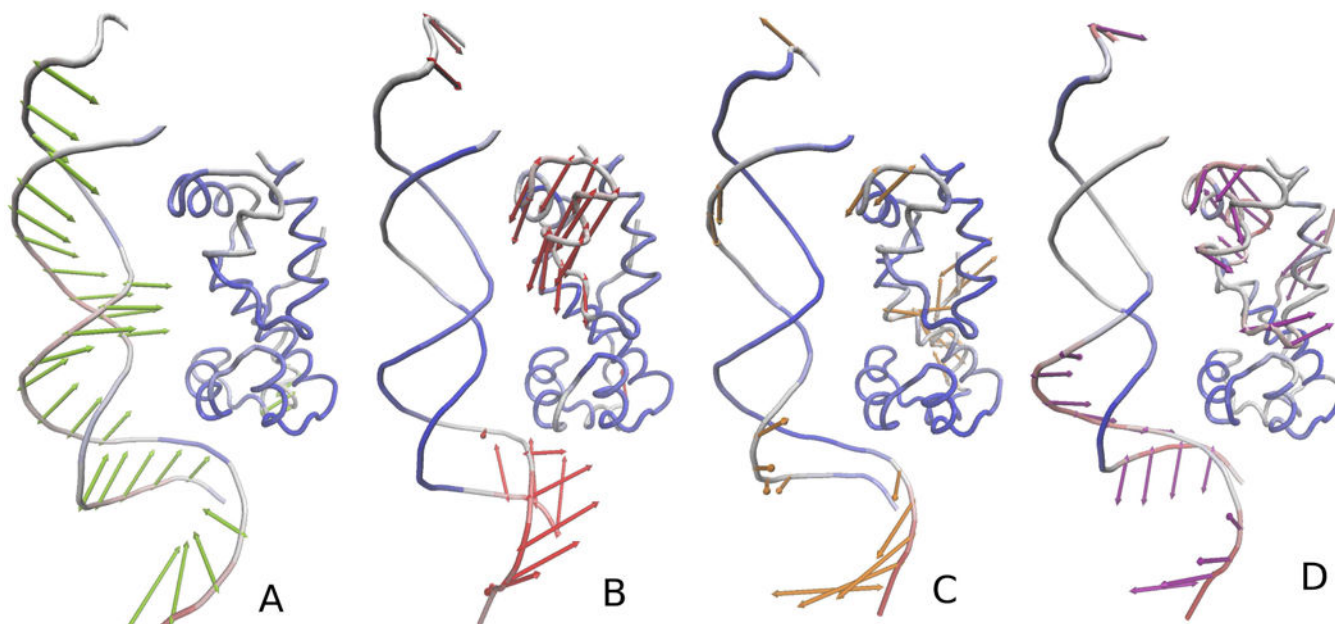
**Figure 7.** Comparison of the crystal structure of multiple antibiotic resistance protein (MarA) bound to DNA (A, PDB code: 1BL0) with the representative structure obtained in the coarse-grained MD simulation (B). Each chain is colored in a rainbow style from blue (for the N-terminus and the C5' end) to red (for the C-terminus and the C3' end).



**Figure 8.** Contact map for MarA DNA interaction (A) crystal structure, (B) last 10 structures from simulations which started from 5 Å separation, (C) last 10 structures from simulations which started from 8 Å separation



**Figure 9.** Plot of the deviations of the C<sup>α</sup> atoms and the C5' atoms of the representative structure of the MarA-DNA system obtained in coarse-grained MD simulations from the experimental structure of MarA (PDB code: 1BL0) as a function of residue index (A) and of the DNA bound to MarA (DNA chain B: red, DNA chain C: black) (B).



**Figure 10.** Illustration of the first four principal components (panels A-D) obtained by doing the PCA of the snapshots of the coarse-grained MD simulations of MarA binding to DNA. The principal components are represented by arrows colored green (A), red (B), orange (C), and blue (D), respectively. The structure displayed in a given panel has been obtained by adding the respective eigenvector coordinates to the site coordinates; the deviations from the average structure are marked by coloring the respective parts of the system from blue (low deviations) to deep red (high deviations).