JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

# Elastic Network Models of Nucleic Acids Flexibility

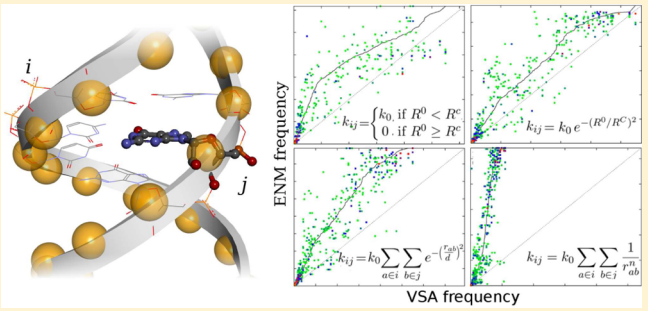
Piotr Setny*,† and Martin Zacharias‡

†Centre for New Technologies, University of Warsaw, 00-927 Warsaw, Poland

‡Physics Department T38, Technical University Munich, 85748 Garching, Germany

Supporting Information

**ABSTRACT:** Elastic network models (ENMs) are a useful tool for describing large scale motions in protein systems. While they are well validated in the context of proteins, relatively little is known about their applicability to nucleic acids, whose different architecture does not necessarily warrant comparable performance. In this study we thoroughly evaluate and optimize the efficiency of popular ENMs for capturing RNA and DNA flexibility. We also introduce two alternative models in which the strength of elastic connections at a coarse-grained level is governed by distance distribution at atomic resolution. For each of the considered ENMs we report the optimal length of spring connections as well as the scaling of elastic force constants that provides the best agreement of vibrational frequencies with normal modes based on atomic force field. In order to determine the absolute values of force constants we introduce a novel method based on the overlap of pseudoinverse of Hessian matrices.




## 1. INTRODUCTION

Nucleic acid molecules play a key role in many biological processes such as gene regulation, splicing, and protein synthesis. Frequently, these processes involve conformational changes in the nucleic acid that are of functional importance during recognition and binding by proteins or for a possible enzymatic activity.[1] Cellular DNA adopts primarily a double-stranded B-form helical structure, and binding of protein molecules can result in global conformational changes such as bending and twisting of the helix as well as local backbone deformations.[2] RNA molecules can form more complex three-dimensional structures that consist only partially of double-stranded (A-form) regions interrupted by mismatched base pairs, bulge structures, or hairpin loop structures that cap the end of helices.[3] Functional requirements can involve conformational changes of all these motifs.

A long-term goal of structural biology is the knowledge of all biomolecular complexes, including all DNA/RNA−protein interactions, in a cell at atomic detail. Their experimental elucidation by X-ray crystallography is currently out of reach due to the large number of putative and transient interactions that are difficult or impossible to crystallize. Computational prediction of the interaction geometry is of increasing importance to obtain at least models of the complex structures.[4−7] Its success is critically dependent, however, on robust models of macromolecular flexibility necessary to guide the transition from unbound to bound geometries of binding partners. In principle, molecular dynamics (MD) simulation techniques are capable of sampling conformational space, treating each degree of freedom at atomic resolution explicitly. However, due to limitations of force field accuracy and computational demand it is neither feasible nor desirable to sample all flexible degrees of freedom during computational evaluation of possible protein−nucleic acid interaction geometries.

Given the unbound structures of putative binding partners, it is possible to precalculate soft, collective degrees of freedom and use these as additional variables during docking of biomolecular complexes.[8] Such soft, collective degrees of freedom can be obtained as principle components of motion from MD simulations[9,10] or by normal-mode analysis of binding partners.[11,12] In particular, normal modes obtained form an elastic network model (ENM) of protein molecules[13] have been shown to overlap often quite well with experimentally observed backbone conformational changes.[14−17] In an ENM the mobility of atoms is controlled by a set of harmonic springs connecting atom pairs, with an equilibrium length that corresponds to the separation in a reference structure and a force constant that either decreases continuously or abruptly as a function of the equilibrium length.[13,15,18] The set of harmonic force constants determines the Hesssian matrix of the ENM which can be diagonalized to rapidly calculate the normal modes and corresponding frequencies. Domain motions in a protein or the differences in backbone structure of a protein in complex with a partner (bound conformation) vs the corresponding unbound structure can often be described by a small set of soft normal modes associated with small eigenvalues (low vibrational frequencies).[14−17] The inclusion of ENM-derived normal modes as

Received: September 16, 2013
Published: November 13, 2013

additional variables offers a computationally efficient approach during systematic docking of proteins and has shown promising results in several cases.[8] It is desirable to extend such an approach to the prediction of putative complex structures formed by nucleic acids and proteins.

Whereas proteins have been studied extensively by a variety of different ENMs,[19] only a limited number of applications to investigate nucleic acid flexibility have been described.[20−25] Protein molecules consist of one or several globular domains, typically folded around a hydrophobic core with little sterically accessible volume inside. In contrast, even folded nucleic acid molecules are much more loosely packed and contain distinct, elongated double helical motifs which are stabilized by specific hydrogen bonds and stacking interactions. Thus, it is not clear if standard ENMs developed for proteins are also optimal in the case of nucleic acids, nor what particular parametrization provides the best results.

The aim of the present study is to systematically evaluate the capability of different ENMs of RNA and DNA molecules to describe experimentally observed conformational changes and nucleic acid mobility. Within the huge spectrum of possible coarse graining schemes, we focus our attention on models that assume a single interaction center per nucleotide, as in general increasing the resolution of coarse-grained representation does not seem to improve the overall performance.[23,26] In addition to considering two models widely used in the context of protein structures (sharp cutoff anisotropic ENM[15] and exponential models[18]), we introduce and test two models in which the strength of spring connections depends on distances at atomic rather than coarse-grained resolution. Such models should have the ability to emphasize base pairing interactions as well as the effect of base flipping out on the stability of a double helix.

The analysis includes the variation of a coarse-grained bead placement within nucleotide monomers as well as the effect of varying the functional formalism of the interaction network. For each considered ENM we determine the strength of force constants providing the best agreement of vibrational frequencies with normal modes obtained with the use of all-atom force field and implicit solvent model. For this purpose we introduce a novel method based on the overlap of pseudoinverse of Hessian matrices. The performance of ENMs is analyzed for a large test set of experimentally determined DNA and RNA structures and by comparison with MD simulations. The consequences and limitations for describing conformational changes in nucleic acids using ENMs are discussed.

## 2. METHODS

**2.1. Elastic Network Models.** Low-frequency vibrations of a molecular system are determined by its general shape and packing density rather than details of local interatomic contacts. This motivates a construction of ENMs in which the system is represented as a set of interaction centers (beads) connected by Hookean springs. Its potential energy is defined as a sum of harmonic terms with equilibrium lengths corresponding to native geometry, $\mathbf{R}^0$, and pairwise specific force constants, $k_{ij}$:

$$U(\mathbf{R}) = \frac{1}{2}\sum_{i,j} k_{ij}(R_{ij}^0)(R_{ij} - R_{ij}^0)^2 \tag{1}$$

where the summation extends over all pairs of beads $i,j$. The optimal number of beads and their placement (coarse-graining scheme) as well as topography and strength of spring connections remain to be determined in the context of specific systems of interest.

In the case of nucleic acids we decided to explore four different models governing the spring topography:

*Sharp Cutoff Model (SC).* All beads whose equilibrium distance falls within a certain cutoff, $R^c$, are connected by uniform springs:[15]

$$k(R^0) = \begin{cases} k_0 & \text{if } R^0 < R^c \\ 0 & \text{if } R^0 \geq R^c \end{cases} \tag{2}$$

*Exponential Model (EX).* Spring constant decays smoothly with increasing distance between beads:[18]

$$k(R^0) = k_0 e^{-(R^0/R^c)^2} \tag{3}$$

*Power Contact Model (PC).* The strength of spring connecting two given beads, $i$ and $j$, depends on pairwise distances of all heavy atoms (in all-atom representation) contributing to their respective beads, according to $r^{-n}$ power model:

$$k_{ij} = k_0 \sum_{a\in i}\sum_{b\in j} \frac{1}{r_{ab}^n} \tag{4}$$

*Exponential Contact Model (EC).* Similar to the PC model, with the only difference that atomic contributions are scaled according to exponential function of pairwise distances:

$$k_{ij} = k_0 \sum_{a\in i}\sum_{b\in j} e^{-(r_{ab}/d)^2} \tag{5}$$

*Random Model.* This model was used for comparison with the above "physically reasonable" ENMs. Here, elastic force constants were randomly assigned from uniform distribution over $[0, 1]$ range.

The optimal bead placement as well as values of control parameters providing the best performance for each of the considered models were determined using a set of test criteria outlined in the following.

**2.2. Comparison Criteria.** In order to select the best possible model and its parametrization we introduced several metrics to evaluate different properties of the produced normal modes directions and frequencies as well as the quality of motions described by EN normal modes.

*Deformation Space Overlap.* We considered a cumulative overlap of 10 lowest-frequency normal modes with the observed deformation space. As reference structures serving for the construction of ENMs we used unbound conformations in the case of bound/unbound pairs of crystallographic structures and models closest to average structure in the case of NMR ensembles or MD simulation trajectories. For pairs of crystallographic structures the deformation space was defined as a single deformation vector describing Cartesian displacements for the transformation from unbound to bound conformations. In the case of NMR ensembles or MD trajectories, deformation space corresponded to a set of 10 most important eigenvectors obtained by principal component analysis (PCA) of structural variants for the given molecule. PCA was carried out with the use of Wordom program.[27] The actual value of deformation space overlap descriptor, $\Omega_{10}$, was calculated as

$$\Omega_{10} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{10} (\mathbf{v}_i\mathbf{u}_j)^2 \tag{6}$$


dx.doi.org/10.1021/ct400814n | J. Chem. Theory Comput. 2013, 9, 5460−5470

where $N = 1$ or $N = 10$ in the case of X-ray and NMR/MD sets respectively, $\mathbf{v}$ are normalized vectors describing the observed deformation space, and $\mathbf{u}$ are EN normal modes. Both in the case of X-ray and NMR/MD-based descriptors, $\Omega_{10}$ values scale from 0 to 1 with increasing space overlap, but due to different dimensions of the deformation space in both cases, values for the two sets are not directly comparable.

*Structural Deformations.* An important application of EN normal modes is their use in molecular docking studies to efficiently guide the transformation between unbound and bound geometries of the binding partners. In principle, the set of all $3N - 6$ vibrational normal mode vectors forms a complete basis for describing any arbitrary deformation (with no overall translation or rotation) of a structure composed of $N$ beads. In the case of macromolecular docking, ENMs aim at the description of global deformations, and thus, their efficiency can be quantified by the minimal number of low-frequency normal modes necessary to describe a conformational transition with desired accuracy.

Here, we consider a descriptor, $f_{0.5}$, indicating the fraction of all determined vibrational normal modes necessary to account for structural deformation that covers 0.5 of the root-mean-square deviation (RMSD) between initial and final geometries. For two structural variants A and B, whose conformations are described by $\mathbf{R}_A$ and $\mathbf{R}_B$ vectors, respectively, and a set of $3N - 6$ vibrational normal modes $\{\mathbf{u}\}$ for the structure A, a transformation of the structure A into B can be gradually performed by using an increasing number of $n$ basis vectors: $\mathbf{R}^n_{A'} = \mathbf{R}_A + \sum_{i=1}^{n}(\mathbf{u}_i\mathbf{R}_{AB})\mathbf{u}_i$. Given that, we define $f_{0.5}$ as

$$f_{0.5} = \frac{\min(n)}{3N-6}: \mathrm{RMSD}(\mathbf{R}^n_{A'}, \mathbf{R}_B) \leq 0.5\mathrm{RMSD}(\mathbf{R}_A, \mathbf{R}_B) \tag{7}$$

Deformations of unbound into bound conformation were considered for the pairs of X-ray structures, while in the case of NMR ensembles, an average value of $f_{0.5}$ for deformations of the closest to average model into all the remaining models was calculated for each structure.

*Fluctuation Amplitudes.* The ability to predict the amplitude of local fluctuations was evaluated by comparing predictions based on harmonic analysis with root-mean-square fluctuations (RMSF) of bead positions determined from NMR ensembles or MD trajectories. The amplitude of positional fluctuations for a bead $i$ can be obtained from from normal-mode analysis by[11]

$$\langle \Delta r_i^2 \rangle^{1/2} = \left[ \frac{k_B T}{m_i} \sum_k \frac{|\mathbf{u}_{ki}|^2}{\omega_k^2} \right]^{1/2} \tag{8}$$

where $\mathbf{u}_{ki}$ is a three-dimensional component of vibrational eigenvector $\mathbf{u}_k$, describing Cartesian displacement of bead $i$, $m_i$ is a bead mass, $\omega_k$ is the frequency of $k$-th normal mode, and $k_B T$ is the Boltzmann constant times temperature. As a measure of agreement between fluctuations derived from NMR ensemble or MD trajectory for a given structure and fluctuations predicted by ENM we used the Pearson's correlation coefficient, referred to in the following as $R_B$.

*Structure Preservation.* In order to test which of the proposed flexibility models provides the best range of structurally sound motions in five lowest-frequency normal modes we considered stepwise deformations of all-atom structures projected on normal modes directions. Starting from the energy minimum (equilibrium structure), atoms used to define beads in coarse-grained representation were shifted in the given normal mode direction in steps that increased RMSD with respect to original geometry by 0.5 Å. After each such shift their positions were restrained, and potential energy of the structure was minimized in all-atom force field. The ff12SB Amber force field with generalized Born (GB) implicit solvent model was used for this purpose, and potential energy minimization was carried out using Nucleic Acid Builder (NAB) from AmberTools package.[28] Finally, base pairing for helical regions was evaluated with the use of X3DNA program.[29] Such procedure was iterated until breakage of hydrogen bonds between paired bases was detected, and the range of deformation, $\Delta_i$ (a sum of ranges for deformations in two opposite normal-mode directions) was recorded for each normal mode $i \in \{1, ..., 5\}$. An average amplitude, $A$, of disruption-free motion for each considered structure was than evaluated as a geometric mean (the product of amplitudes of orthogonal modes corresponds to a "volume" of disruption-free deformation space, and their geometric mean corresponds to an edge of a hyper cube of equal volume) over five considered normal modes:

$$A = \frac{1}{2}\left(\prod_{i=1}^{5} \Delta_i\right)^{1/5} \tag{9}$$

**2.3. Vibrational Frequencies and Force Constants.** In order to obtain a physically meaningful representation of molecular elasticity by the considered ENMs, we scaled force constants ($k_0$ values in eqs 2−5) to achieve the best agreement of vibrational frequencies with a reference model based on normal-mode analysis in all-atom force field representation.

It should be noted that ENMs are expected to perform reasonably rather in low-frequency limit of molecular vibrations, and further, that the comparison of frequencies is only meaningful between vibrational modes of similar directions. A convenient way of obtaining the best $k_0$ values is based on Hessian matching technique. Once noting that a Hessian matrix derived from ENM can be represented as $H_{ENM} = k_0\tilde{H}_{ENM}$, where $\tilde{H}_{ENM}$ carries all geometry-dependent information, and $k_0$ plays a role of a scaling constant, one can look for $k_0$ which maximizes an overlap between $H_{ENM}$ and a reference Hessian derived from all-atom force field model, for instance, by applying vibrational subspace analysis (VSA) procedure[30,31] (see below), in order to match the dimensionality of $H_{ENM}$. An overlap between two mass-weighted Hessian matrices $A$ and $B$ can be assessed by a norm based on their pseudoinverse representation:[32,33]

$$s(A, B) = 1 - \frac{\sqrt{\sum_i (\omega_i^A)^{-2} + \sum_j (\omega_j^B)^{-2} - 2\sum_{ij} (\omega_i^A \omega_j^B)^{-1} (\mathbf{u}_i^A \mathbf{u}_j^B)^2}}{\sqrt{\sum_i (\omega_i^A)^{-2} + \sum_j (\omega_j^B)^{-2}}} \tag{10}$$

where $\omega_i$ and $\mathbf{u}_i$ are vibrational frequencies (square roots of eigenvalues) and eigenvectors, respectively. This measure is 0 for completely orthogonal matrices and 1 for identical matrices. The result depends on the degree of collinearity of eigenvectors, which for each pair is weighted by frequency values. The use of inverted frequencies implies that the lowest frequencies contribute most to the similarity, which coincides well with the region of frequency spectrum where ENM is expected provide meaningful results. Given the above, the task of finding optimal scaling for elastic forces reduces to finding $k_0$ such that $s(k_0\tilde{H}^{MW}_{ENM}, H^{MW}_{VSA}) = \max$, which can be easily

performed using Newton optimization method (a MW superscript indicates mass-weighted Hessian matrices). The final $k_0$ value for each ENM model, was obtained as an arithmetic average over all structures considered for parametrization.

A coarse-grained mass unweighted Hessian, obtained with the use of VSA of all-atom force field based Hessian, was also used to determine the dependence of (pseudo)elastic force constants on the distance between beads. By isolating from VSA Hessian matrix a 3 × 3 off-diagonal superblock $\mathbf{K}_{ij}$ that gathers second derivatives of potential energy with respect to Cartesian coordinates of beads $i$ and $j$, one can calculate

$$k_{ij} = -\mathrm{Tr}(\mathbf{K}_{ij}) \tag{11}$$

which corresponds to harmonic force constant, under assumption that bead−bead interaction is described by a harmonic potential $V_{ij} = 1/2k_{ij}(\Delta r_{ij})^2$.[34]

All-atom Hessians were obtained with the use of NAB tool. Structural models of interest were supplemented with hydrogen atoms, and their potential energies were thoroughly minimized in ff12SB force field and GB solvent model (with default NAB settings) in order to ensure that all vibrational frequencies were positive. Hessian matrices were then extracted and subjected to VSA procedure. Briefly, in the course of VSA all-atom Hessian $H$, and mass matrix $M$ are rearranged into sub-blocks describing the subsystem of interest $s$ and environment $e$:

$$H = \begin{pmatrix} H_{ss} H_{se} \\ H_{es} H_{ee} \end{pmatrix}, \quad M = \begin{pmatrix} M_{ss} M_{se} \\ M_{es} M_{ee} \end{pmatrix} \tag{12}$$

Here, the subsystem of interest corresponded to coarse-grained beads located at geometric centers of the ribose rings whose positions were defined as linear combination of ring atom coordinates. The environment corresponded to the rest of the structure. An effective Hessian, describing interactions in the coarse-grained system, is obtained as[33] $H_{\mathrm{VSA}} = H_{ss} - H_{se}(H_{ee})^{-1}H_{es}$. Accordingly, an effective mass matrix is obtained as $M_{\mathrm{VSA}} = M_{ss} + H_{se}(H_{ee})^{-1}M_{ee}(H_{ee})^{-1}H_{es} - H_{se}(H_{ee})^{-1}M_{es} - M_{se}(H_{ee})^{-1}H_{es}$. The corresponding normal-mode equation:

$$H_{\mathrm{VSA}}\tilde{\mathbf{u}} = \omega^2 M_{\mathrm{VSA}}\tilde{\mathbf{u}} \tag{13}$$

can be then solved in a standard manner, by introducing a mass-weighted Hessian $H_{\mathrm{VSA}}^{\mathrm{MW}} = M_{\mathrm{VSA}}^{-1/2}H_{\mathrm{VSA}}M_{\mathrm{VSA}}^{-1/2}$ and obtaining mass-weighted normal modes $\mathbf{u}$.

**2.4. Data Sets.** *Structures.* We used four sets of nucleic acids structures taken from the Protein Data Bank (PDB)[35] for ENM parametrization and evaluation of their performance: (A) 77 NMR structures representing pure RNA molecules (without small ligands and proteins), each of which contained at least 20 structural models, (B) 7 pairs of unbound and protein-bound RNA crystallographic structures, (C) 45 NMR structures of pure DNA molecules, each containing at least 10 structural models, and (D) 41 pairs of protein-bound helical DNA crystallographic structures and corresponding DNA models of identical nucleotide sequence and ideal B-DNA form taken from protein−DNA docking benchmark.[36] A detailed list of all structures is given in Supporting Information (SI).

In addition, we used two long MD simulations of RNA and DNA double helices in explicit solvent. Technical details of MD simulations are given in SI.

*Structural Descriptors.* In order to characterize structures used in the study and their deformations we used several descriptors. Among structural features we considered size (number of nucleotides), RMSD, helix content (the ratio of nucleotides involved in helical base pairs to the total number of nucleotides), and normalized radius of gyration, $nR_{\mathrm{gyr}} = N^{-1/3}R_{\mathrm{gyr}}$, where $N$ is the number of beads in the structure, and $N^{1/3}$ is proportional to the radius of spherical assembly of the same number of beads (uniform packing density, i.e., uniform volume per bead is assumed for all structures). In order to assess the collectivity of deformations, we evaluated their entropies as: $S = -\sum_i^N(\Delta r_i)^2 \ln(\Delta r_i)^2/\ln N$,[33] where $\Delta r_i$ are positional displacements of individual beads. Here, for theoretically most collective motions, i.e., if each bead displacement equals to $1/\sqrt{N}$ (note that displacement vectors are normalized: $\Sigma_i^N(\Delta r_i)^2 = 1$), $S = 1$, and for most localized modes, where only one bead moves, $S = 0$.

## 3. RESULTS

**3.1. Characterization of the Data Sets.** The four considered data sets contain together 170 structures of nucleic acids with sizes ranging from 10 to over 100 nucleotides. In general, both X-ray sets gather structures that are larger than those in NMR sets, with three tRNA structures and one 7SL RNA being the largest. As evidenced by typically larger normalized radii of gyration, X-ray structures tend to be also more elongated than NMR structures, in particular in the X-ray DNA set, where all structures adopt straight, rod-like B-DNA geometry.

A double helical arrangement is the most common secondary structure motif: on average roughly 85% of nucleotides in each structure takes part in double helices in all sets except for X-ray DNA where, due to the fact that unbound structures were deliberately chosen as ideal B-type DNA, all nucleotides are involved in canonical base pairs. The regularity of the remaining helical structures varies considerably. In the most diverse NMR DNA set, only 30% of identified helices can be classified as B-type (according to X3DNA program) and 15% of the structures form triple or quadruple helices. In the NMR RNA set, a helical stem region capped by a loop comprising 4−6 nucleotides is the most common architecture, with on average 60% of helical regions corresponding to A-type RNA. More complex structures such as pseudoknots constitute 6% of the set. In the X-ray RNA set, helical regions are more regular, with over 70% corresponding to A-type geometry (see Figure 1).

The amplitude of conformational changes is typically larger for X-ray sets than NMR sets, which can be a consequence of the fact that NMR ensembles reflect natural flexibility of the considered molecules, whereas transitions between unbound and bound states are promoted by binding partners. The collectivity of motions, as measured by the entropy of distribution of atomic displacement amplitudes, is quite similar for both NMR sets, while the X-ray sets show contrasting behavior. In the DNA set there is a clear tendency for highly collective deformations, whereas in the RNA set there is a high proportion of structures exhibiting rather local (low entropy) deformations. Indeed, the analysis of complexes formed with protein partners reveals that the in the case of DNA, the most frequently observed binding mode is achieved by a bend of helical axis with no deformation in local helix geometry. In turn, in the case of low-entropy deformations in RNA structures binding involves local distortions of nucleic acid backbone.

**3.2. Optimal Bead Placement.** Prior to comparing different ENMs, for each model we systematically tested the effect of bead positioning within the nucleic acid backbone on the efficiency of the resulting normal modes in representing

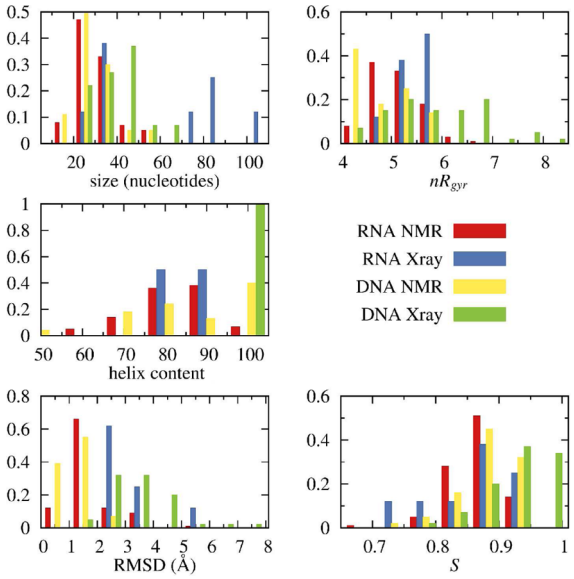



**Figure 1.** The distribution of features among four considered data sets.

global molecular deformations. Using the sets of RNA and DNA NMR ensembles, we evaluated such efficiency based on the overlap between the first 10 EN normal modes with the first 10 principal components of NMR ensembles ($\Omega_{10}$ parameter).

The obtained results (Figure 2) indicate that both in the case of RNA and DNA structures ENMs perform more efficiently with beads placed within the ribose ring rather than on the phosphorus atom, as is often assumed by default.[20,22−25] It most likely reflects the importance of maintaining relatively strong connections between nucleotides that form pairs in double helices.

Bead placement at the geometric center of the ribose ring (pseudoatom B in Figure 2) systematically occurs within the two best positions (as the best one or insignificantly behind the best: 0.98 vs 1.00 normalized score) in all models for RNA structures or as the best or the third one (in SC and EX models) for DNA structures. A qualitatively identical trend, though with slightly different values of optimal control parameters (see below), was observed for the sets of crystallographic structures, and therefore, bead placement at the center of ribose ring was adopted as universal and selected for all further comparisons.

**3.3. Optimal Control Parameters.** The lower bound for the length of spring connections is imposed by the need to avoid splitting the system into insufficiently connected subnetworks that can freely rotate or translate with respect to each other. In the case of SC model, cutoff length below 14 Å (eq 2) led to artifacts for several structures, as did the cutoff of $R^c$ < 4 Å (eq 3) for EX model. In the case of both contact models, the length of effective spring connections depends not on direct bead−bead distance but on the spatial arrangement of heavy atoms belonging to the respective nucleotides. Here, properly determined elastic networks were obtained for all the considered structures for $n \leq 10$ (eq 4) and $d \geq 4$ Å (eq 5) for PC and EC models, respectively.

Beyond the shortest reasonable length scale, the performance of ENMs appears to vary slowly with changes in control parameters (Figure 3). In most cases, possibly optimal efficiency has a tendency to extend as a plateau region rather than to peak around a certain value. Moreover, the performance evaluated with the use of control parameters optimized

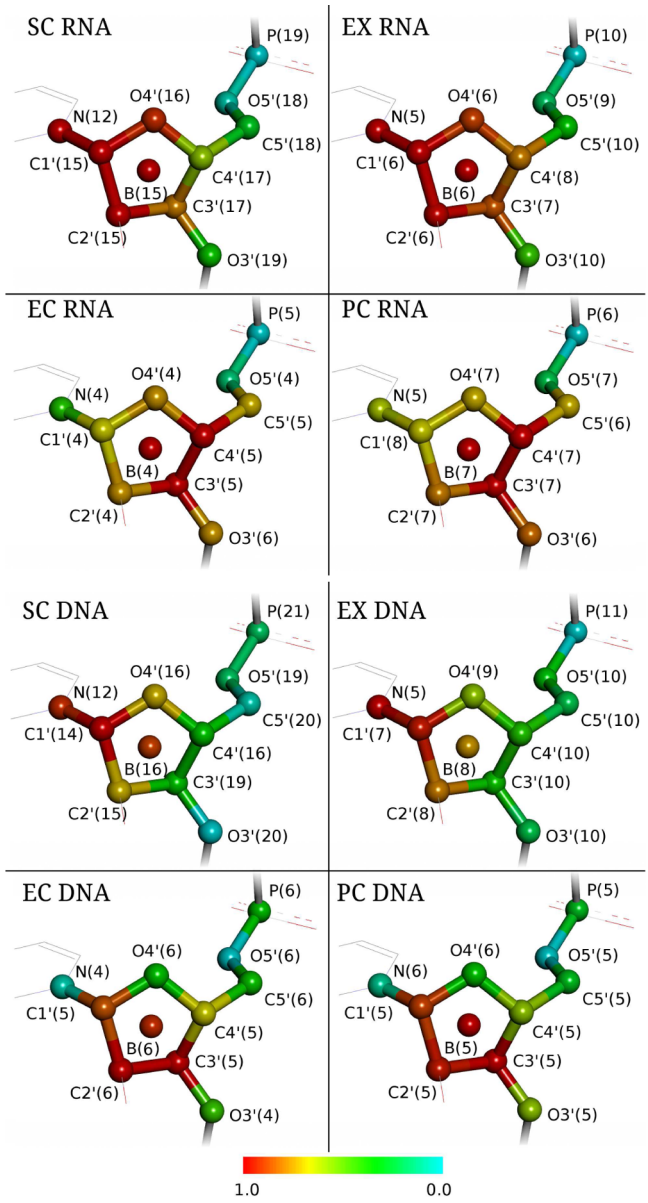



**Figure 2.** The influence of bead placement on the performance for RNA and DNA NMR sets evaluated as overlap between 10 lowest frequency ENM normal modes and 10 most important PCA modes ($\Omega_{10}$ descriptor). The results are rescaled for each model to fit between 1.0 (best performance) and 0.0 (worst performance).

*individually* for each structure (solid lines in Figure 3) is only marginally better than the performance evaluated with uniform values of control parameters for the entire sets, which indicates that there is very little room for improvement due to structure-dependent adjustment of control parameters.

Indeed, parameters optimized for individual structures display little correlation with global structural descriptors (Figure 4). Only limited positive correlation ($|R| \lesssim 0.3$) was found between structure size (the number of beads) and elongation (normalized radius of gyration) and spring length providing optimal results in $f_{0.5}$ and $R_B$ tests, while no such correlation at all was found for $\Omega_{10}$ results. Interestingly, in the latter case a small ($|R| \lesssim 0.3$) but clearly negative correlation was found with the amount of double helical geometry. A small negative correlation was also found between the optimal spring length and the collectivity (entropy) of deformation to be


dx.doi.org/10.1021/ct400814n | J. Chem. Theory Comput. 2013, 9, 5460−5470

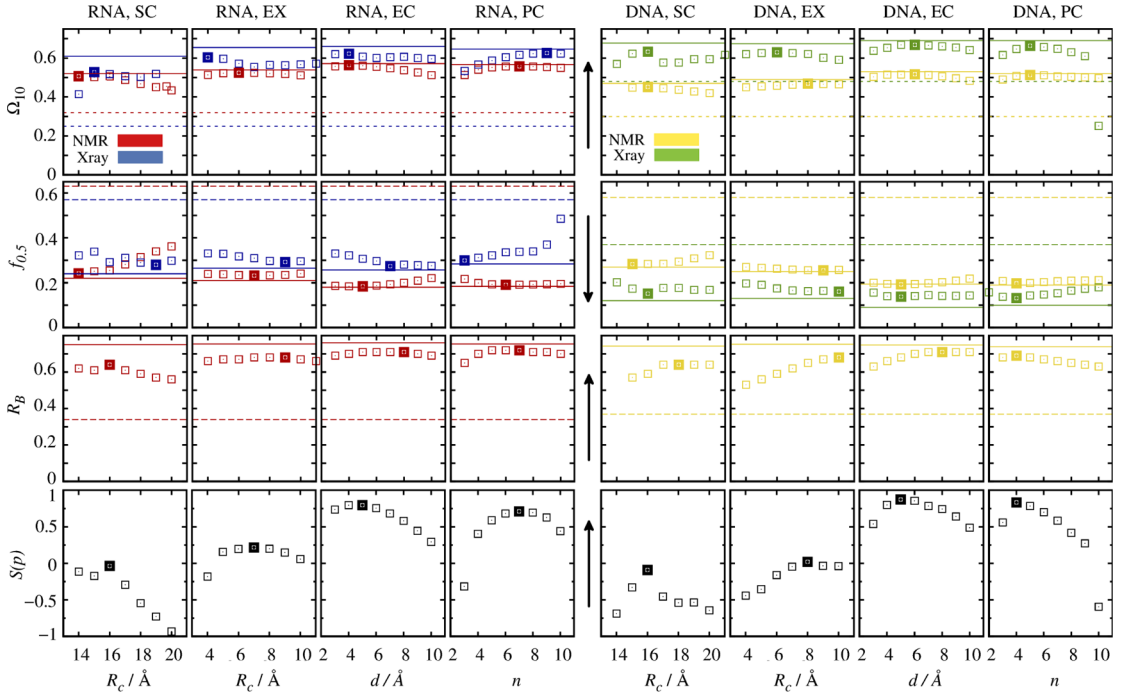



**Figure 3.** Performance of the models for different values of control parameters (eqs 2−5). Bold symbols indicate best average performance for a given value of control parameter; solid lines indicate average performance for individually optimized control parameters; and dashed lines indicate performance for random ENM model. $\Omega_{10}$ indicate deformation space overlap; $f_{0.5}$ indicate fraction of normal modes necessary to cover the desired deformation; $R_B$ indicate correlation of predicted fluctuation amplitudes with NMR mobilities; and $S(p)$ indicate normalized score. Arrows indicate the desired behavior of respective target functions.

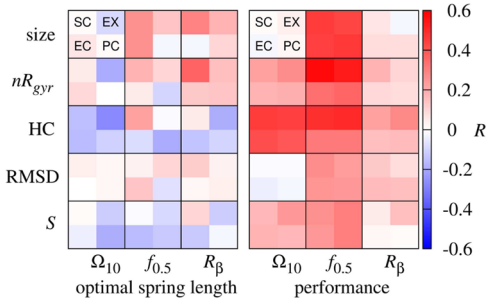



**Figure 4.** Correlations (expressed as Pearson's correlation coefficients) between different features of the considered structures and optimal length of spring connections (left) and ENM performance (right). The length of spring connections was defined as a minimal distance $R_0$ at which $k(R_0)/k_0 < 0.01$.

explained by the model. Together those two findings indicate that global, collective deformations of regular, helical structures tend to be better reproduced by ENMs with possibly short spring connections. Nonetheless, given relatively insignificant correlations, it would be hard to gain from this observation any predictive value that would be useful for the optimization of ENM in the context of individual structures.

Another moderate but clearly visible trend in ENM behavior is a shift to longer optimal spring connections in tests involving an increasing number of elastic normal modes (Figure 3): from $\Omega_{10}$ (10 lowest-frequency modes) to $f_{0.5}$ (50% of all modes) to $R_B$ (all modes). A similar observation has been reported for protein structures.[37] Such effect may be explained by the fact that the density of modes tends to shift toward higher frequencies as the length of spring connections increases,[15] hence changing as well the length scale of best reproduced deformations.

In general, the lack of definite correlation between structural features and the optimal length of spring connections together with relatively flat dependence of the general performance on control parameters seem to justify the use of only single values of control parameters in each type of ENM model. In order to determine parameters that provide possibly universally good results, separately in the case of RNA and DNA structures, we constructed a consensus score, $S(p)$, which reflects a combined performance in covering global deformations for NMR and X-ray structures together with predicting fluctuation amplitudes in agreement with NMR ensembles:

$$S(p) = w_1 \frac{\tilde{\Omega}_{10}(p) + \tilde{f}_{0.5}(p)}{2}\Bigg|_{\text{NMR}} + w_1 \tilde{R}_B(p) + w_2 \frac{\tilde{\Omega}_{10}(p) + \tilde{f}_{0.5}(p)}{2}\Bigg|_{\text{X-ray}} \quad (14)$$

Here, $p$ is a value of a control parameter, $\tilde{\Omega}_{10}$, $\tilde{f}_{0.5}$, and $\tilde{R}_B$ are results of the respective tests standardized according to mean value and standard deviation of data obtained for a given test for all considered ENMs, $w_1$ and $w_2$ are weights proportional to the number of NMR and X-ray structures: $w_1/w_2 = 10.0$ in the case of RNA and 1.1 in the case of DNA structures.

Even though data for DNA and RNA structures were obtained and processed totally independently, the resulting consensus scores, and in particular optimal values of control parameters, are almost identical for both kinds of nucleic acids with one exception of the PC model (Figure 3). As might have been expected, the SC model appeared to be the most sensitive to the length of spring connections, with cutoff distance of 16 Å being clearly optimal both in the case of RNA and DNA. Presumably, relatively sharp decrease in performance for


dx.doi.org/10.1021/ct400814n | J. Chem. Theory Comput. 2013, 9, 5460−5470

shorter as well as longer cutoff distances is related to the regularity and prevalence of helical secondary structure: addition or removal of springs connecting similar positions within repeating elements of a double helix may has a notable effect on the elastic properties of the whole helical domains.

**3.4. Performance of the Models.** *General Descriptors.* The results indicate small but persistent advantage of both contact-based models over standard, sharp cutoff and exponential ENMs. Their consensus performance over almost the entire considered range of control parameters is in most cases still better than the best results for any of the SC or EX models. This advantage may be a result of better reproduction of base-pairing interactions by contact-based models as well as their ability to capture local destabilization of helical structure due to flipped out bases. Nonetheless, variations in performance between the models seem to be on average quite insignificant, confirming observations known from protein modeling endeavors of little sensitivity of ENMs to the details in spring connection schemes. From the general, topological point of view, all four considered ENMs describe an elastic network in which the strength and direction of forces acting on a single node is dictated by density distribution of its neighbors. As evidenced by definitely worse performance of random elastic networks in all considered categories (dashed lines in Figure 3), such topology indeed captures reasonably well mechanic properties of macromolecules,[38] however, consistent, systematic improvements beyond the observed baseline fidelity are hard to achieve.

Indeed, differences in performance between the models are generally smaller than differences in performance between the four data sets, indicating that efficiency of ENMs depends more on individual properties of a given structure rather than the assumed topography of elastic contacts. Expectedly, we found that the performance for all considered ENMs, in particular with respect to the reproduction of global flexibility measured by $\Omega_{10}$ and $f_{0.5}$ descriptors, is better for elongated molecules (positive correlation with normalized radius of gyration, $nR_{gyr}$) with a large proportion of double helical secondary structure (positive correlation with helical content, HC). Large contribution to global deformations originates in such cases from bending of the helical axis, a type of motion that should be efficiently represented by any reasonable elastic model of an elongated structure. Also expectedly, low-frequency EN normal modes were found to provide better overlap with deformations of more collective nature, as evidenced by a positive correlation of performance measured by $\Omega_{10}$ and $f_{0.5}$ descriptors with deformation entropy ($S$).

In contrast to the results for $\Omega_{10}$ descriptor, ENMs performance evaluated by $f_{0.5}$ descriptor shows positive correlation with structure size and RMSD of conformational change. This can be understood by taking into account that $\Omega_{10}$ values depend only on the overlap of deformation *directions*, while $f_{0.5}$ values are also sensitive to their *amplitudes*. In the case of large structures with large RMSD, explaining 50% of the initial deformation typically requires global shape modification which can be efficiently described by low-frequency EN normal modes, while leaving difficult to represent, residual local deformations within the remaining 50% of the initial RMSD. In contrast, initially small RMSD often arises from only local shape variations, and reducing it by 50% already requires normal modes of higher frequency.

Among the four data sets, overall best results were obtained for X-ray DNA set, whose all members have regular double helical, rod-like structure and whose deformations are most collective and have high amplitudes (see Section 3.1 and Figure 1). Notably, even the results of random ENM for X-ray DNA set were not much worse than the results of "physically reasonable" ENMs for other sets. The difference in performance between X-ray DNA and X-ray RNA sets is definitely more pronounced for $f_{0.5}$ descriptor in comparison to $\Omega_{10}$. Based on inspection of bound/unbound partners, one possible explanation here is that mutual RMSD in the case of RNA structures has definitely more contributions due to local structural rearrangements (including folding or unfolding of secondary structure) which, as discussed above, require generally higher frequency normal modes to be reproduced.

Among ENMs for the two NMR sets, consistently worse results in categories related to global flexibility ($\Omega_{10}$ and $f_{0.5}$ descriptors) were obtained for DNA structures, which likely reflects larger structural irregularity of the DNA set in comparison to the RNA set (see Section 3.1). On the contrary, practically no differences were found with respect to local fluctuations ($R_B$ descriptor), in accordance with small observed correlation of $R_B$ performance with descriptors referring to structural features (Figure 4).

*Comparison with MD.* All four ENMs performed similarly in the reproduction of dynamical properties (first 10 principal modes, and positional fluctuations of coarse-grained centers) for single RNA and DNA helical structures, derived from long MD simulations in explicit solvent (Figure 5). Interestingly, the

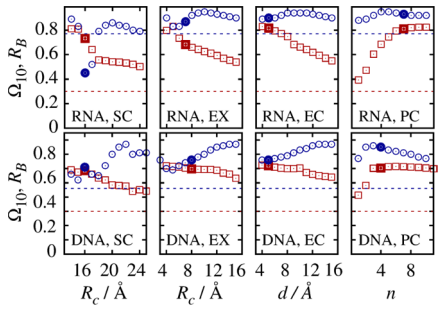



**Figure 5.** The performance of ENMs in reproducing dynamical features determined from MD simulations. Red squares: overlap between 10 lowest-frequency ENM normal modes and 10 most important PCA modes ($\Omega_{10}$); blue circles: correlation between RMSF of coarse-grained beads from MD and their fluctuation amplitudes calculated from ENMs ($R_B$). Bold symbols denote results for optimal control parameters determined by the analysis of NMR and X-ray structures. Dashed lines indicate the performance of random ENM.

results obtained for individually optimal control parameters are nearly the same across all models, but the two contact-based models show somewhat better performance for consensus parameter values determined in the previous section (Figure 5, thick symbols). It suggests that the slight advantage of the contact based models described above for X-ray and NMR sets may be not necessarily a result of better topology of their spring connections but rather a consequence of wider and less structure dependent range of control parameters providing nearly optimal performance.

A clear preference for shorter spring connections for the reproduction of global motions ($\Omega_{10}$ descriptor) when compared with with longer spring lengths providing good reproduction of fluctuations ($R_B$ descriptor) is clearly visible in all cases. As both structures are regular helices (secondary


dx.doi.org/10.1021/ct400814n | J. Chem. Theory Comput. 2013, 9, 5460−5470

structure was preserved throughout entire simulations) and exhibit unrestrained, spontaneous motions during simulations, it seems to confirm the observation based on the analysis of X-ray and NMR structures regarding the tendency to adopt possibly short elastic contacts for the reproduction of collective motions in regular helical structures.

In comparison with results obtained for X-ray and NMR structures, the performance of ENMs in capturing dynamical features determined from MD simulations was exceptionally good, with the overlap of global deformational space in the range of 0.6−0.8 and correlations in amplitudes of local fluctuations reaching 0.95. Interestingly, the results for random ENMs (dashed lines in Figure 5) were also exceptionally good in the case of $R_B$ test (up to 0.78 in the case of RNA) but remained at the same level for $\Omega_{10}$ descriptor (∼ 0.3) as for X-ray and NMR structures.

*Structure Preservation.* In principle, normal modes describe physical motions of the system strictly in the close vicinity of the considered energy minimum. In applications such as molecular docking, it is desirable, however, to use low-frequency normal modes to guide deformations of larger amplitudes. A relevant question, especially in the case of normal modes represented in Cartesian coordinates, is to what extent a structure can be deformed without affecting its physicochemical integrity. In the case of nucleic acids, the first to suffer from disruption is usually the geometry of a double helix. Using methodology outlined in the Methods section we evaluated average amplitudes of motions that preserve base pairing in helical regions (Figure 6). The results are based on six unique unbound structures from RNA-X-ray set and eight randomly selected DNA models from DNA-X-ray set.

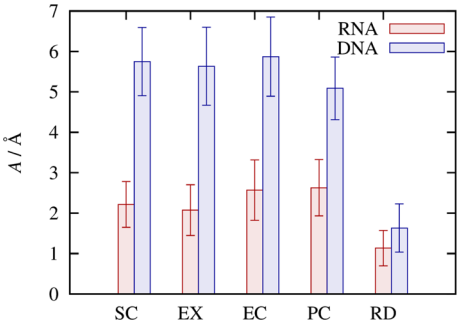



**Figure 6.** The amplitudes of disruption-free motions (eq 9) obtained for the considered ENMs. Error bars correspond to one standard deviation in obtained results.

Again, similarly to previous observations, the differences between four ENMs are relatively small: both in the case of RNA and DNA structures they fall well within respective error bars denoting one standard deviation of the obtained results. Even though contact-based models seem to be better suited for the preservation of helical geometry, as they provide relatively strong interactions between beads corresponding to base-paired nucleotides, they did not perform definitely better than standard SC and EX models, and moreover, the PC model gave the worst results in the case of DNA structures. It indicates that the range of structure preserving deformations determined by elastic models depends more on individual properties of a given structure rather than on the strength and length of spring connections, at least as long as they correspond to a physically reasonable model (note that the observed performance of random elastic network is clearly inferior).

On average, the amplitudes of disruption-free deformations were much larger in the case of DNA structures (∼5.5 Å) than in the case of RNA (∼2.5 Å). Most likely it is due to the fact that DNA structures under study corresponded to ideal B helical conformations with perfectly formed, complementary base pairs, while RNA geometries were based on realistic, imperfect crystallographic structures.

*Vibrational Frequencies.* All properties considered so far were independent of the absolute values of multiplicative parameters $k_0$ that are responsible for scaling the strength of spring connections in each ENM. In principle, proper setting of $k_0$ should provides physically reasonable vibrational frequencies and amplitudes of elastic normal modes. One possible method for determining $k_0$ value is based on the comparison of the amplitudes of theoretically predicted positional fluctuations (eq 8) with the crystallographic B-factors.[39] Here we introduce an alternative, systematic approach relying on the comparison of pseudoinverse of Hessian matrix derived from ENM with pseudoinverse of the reference matrix obtained by a vibrational subspace analysis of full-atom Hessian constructed from atomistic force field model (eq 10). Such a method is sensitive to the agreement in frequency between modes that have similar directions and assigns higher weights to the lowest-frequency modes, thus emphasizing the region of frequency spectrum where ENMs are expected to provide meaningful results.

The obtained average force constants expectedly decrease with increasing length of spring connections in each model (Figure 7). The value of ∼3 kcal/(mol Å$^2$) obtained for the

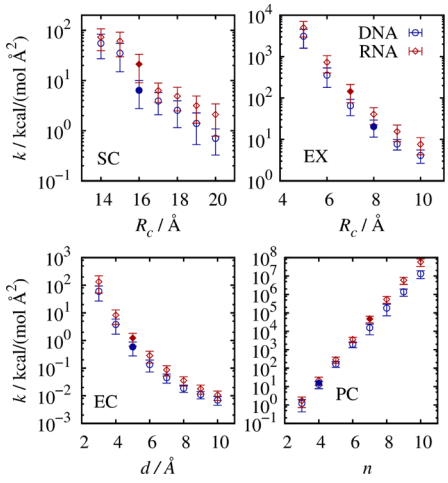



**Figure 7.** Absolute spring constant ($k_0$) values determined for different values of control parameters. Error bars correspond to one standard deviation in obtained results.

cutoff length of 19 Å in SC model for RNA is 20 times larger than the value reported by Bahar for RNA Gaussian network model (GNM) with the same cutoff but interaction centers positioned on phosphorus atoms.[20] Similar ratio of the force constants corresponding to anisotropic network model (i.e., SC model considered here) and GNM was also observed for protein structures.[15]

Force constants for RNA are consistently nearly twice as large as those for DNA, confirming the general view of greater stiffness of RNA in comparison to DNA.[40,41] Certainly, particular values of force constants calculated here are a very


dx.doi.org/10.1021/ct400814n | J. Chem. Theory Comput. 2013, 9, 5460−5470

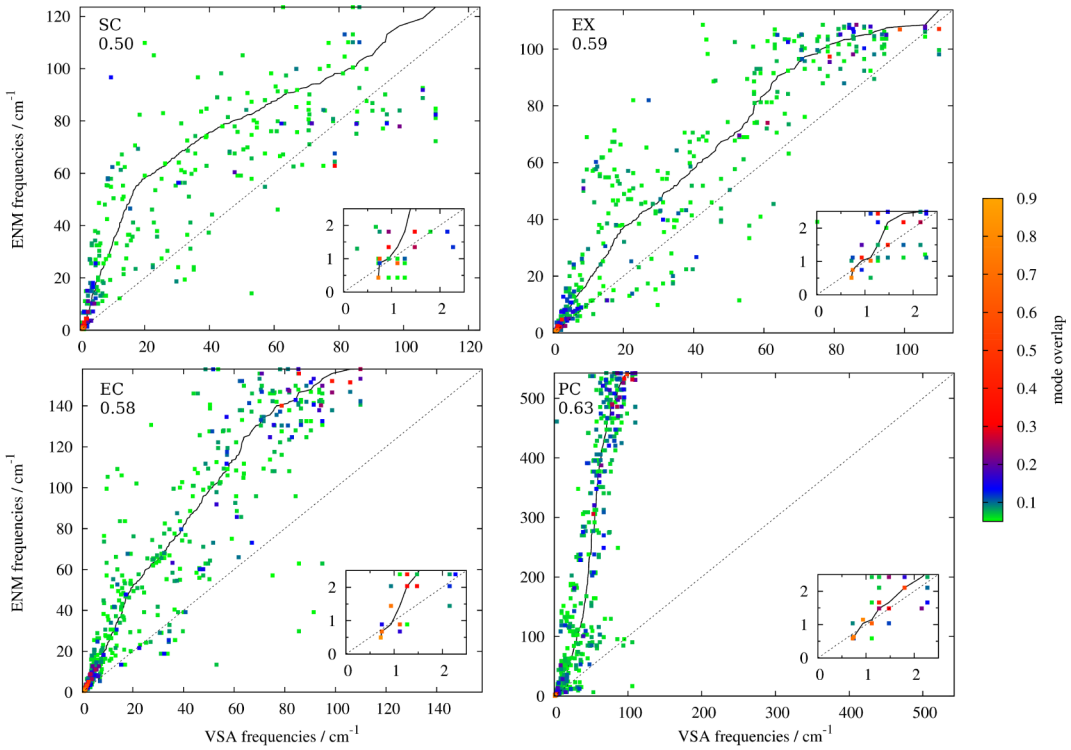



**Figure 8.** The comparison of vibrational frequencies for VSA and ENMs normal modes that have at least 0.05 square overlap. Solid line presents frequencies of VSA vs ENMs consecutive normal modes. Data for aspartyl-tRNA (PDB: 2TRA) are used as an representative example.

general, force field dependent measure of structural rigidity. Still, however, their obtained ratio remains in qualitative agreement with assessments based on force field independent analysis of structural databases[40] as well as on MD simulations.[41]

The comparison of vibrational frequencies obtained for one representative structure (aspartyl-tRNA, PDB: 2TRA) for the considered ENM models with frequencies resulting from VSA of force field based, all-atom Hessian is shown in Figure 8. Apparently, none of the ENMs is able to consistently reproduce the whole range of force field based frequency distribution. As there is no linear transformation between ENM and VSA frequency distribution, an agreement in one region of the spectrum implies discrepancy in another. The weighting scheme used in pseudoinverse Hessian matching procedure, that favors agreement in frequencies and directions of normal modes in the low-frequency region (eq 10), in all cases leads to a shift of medium- to high-frequency region toward higher values. This effect is particularly strong in the PC model. Here the lowest-frequency region is very well reproduced with several consecutive modes showing high overlap with VSA modes and, at the same time, having nearly identical vibrational frequencies. This results in the highest overlap with VSA Hessian among all considered ENMs, but at a price of evidently overestimated higher frequencies. Such behavior can be expected from an elastic model which provides very short and stiff connections. Among all four ENMs, the EX model seems to provide the best balance between the strength of short and long elastic constants leading to a frequency spectrum that is most similarly distributed to the reference one.

In all cases it would be possible to obtain much better agreement of frequencies in medium- and high-frequency regions by using Hessian matrices instead of their pseudoinverse form for the determination of force constants (i.e., using weights proportional to $\omega$ instead of $\omega^{-1}$ in eq 10). The resulting frequency distributions are included in SI. It should be kept in mind, however, that from the applicability point of view the agreement in the lowest-frequency region is most important in tasks such as the estimation of fluctuation amplitudes, vibrational entropies, or global elastic properties.

*Structure of Spring Contacts.* Determination of effective spring constants based on VSA of all-atom Hessian gives insight into relative strength of different classes of coarse-grained interactions. As can be seen in Figure 9, groups of interactions

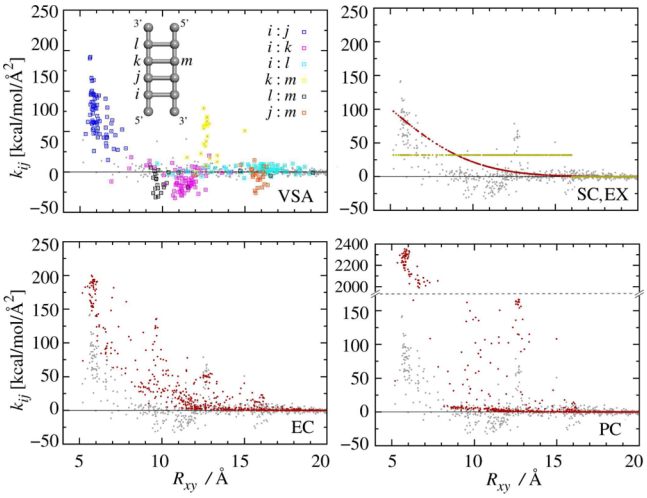



**Figure 9.** The dependence of effective spring constants on bead−bead distance obtained by VSA analysis of all-atom Hessian (gray points) and ENMs (red and yellow points). Upper left plot: dissection of VSA spring constants according to different types of intrahelical contacts. Data based on the analysis of 2TRA PDB structure.


dx.doi.org/10.1021/ct400814n | J. Chem. Theory Comput. 2013, 9, 5460−5470

associated with distinct peaks in effective spring constants correspond to 1−2 and 1−3 interactions between monomers in a double helix. Their length extends roughly to 16 Å, likely justifying the optimal length of spring constants in SC model, determined on the basis of functional tests described above. Interestingly, 1−3 interactions are mostly associated with negative effective spring constants, which indicates their many-body character, impossible to describe with harmonic energy terms. This may pose a limit to the performance of any ENMs of helical nucleic acids and is likely the reason due to which the considered modifications in spring topology resulted only in small variations in overall efficiency of the models.

In agreement with the analysis of frequency distribution, short-range interactions (with equilibrium length <10 Å) tend to be overestimated by ENMs, in particular in contact-based models, giving rise to the observed blueshift in medium- and large-frequency region. In turn, the amplitudes of effective spring constants in the range corresponding to base pair contacts and 1−3 intrahelical angles are in qualitative agreement with VSA model. It suggests the predominant role of those interactions in determining low-frequency motions.

Having insight into the distribution of the strength of coarse-grained interactions, in principle, allows the construction of ENM in which distance-dependent magnitude of spring constants can be tabularized or described by an analytic formula closely following average values from VSA-determined distribution. Among many other ideas, we considered also such model, but we did not observe any particular improvement over four models described in the above (results not presented).

## 4. SUMMARY AND CONCLUSIONS

We have optimized and compared the performance of four different versions of ENMs in reproducing flexibility of RNA and DNA structures. We found that, regardless of particular functional form governing the strength of spring connections, overall better results can be achieved by placing coarse-grained interaction centers within the ribose ring, rather than on phosphorus atom, as is often assumed by default.[20,22−25] Somewhat surprisingly, but in agreement with earlier observations for proteins,[26] variations in spring formalism appeared to have little influence on the performance of ENMs. Furthermore, the performance was not particularly sensitive to changes in control parameters defining effective lengths of spring connections for each model. Accordingly, little or no correlation was found between optimal spring length and structural features such as size or elongation.

Small differences in average performance of the four considered ENMs hamper their quantitative comparison in absolute efficiency scale. In order to finally rank the models we assigned +1 or −1 to each model for the best or the worst performance, respectively, in all tests considered above (i.e., $\Omega_{10}$ for NMR, X-ray and MD sets, $f_{0.5}$ for X-ray and NMR sets, $R_B$ for NMR and MD sets, and $A$ for X-ray sets). The results, summarized in Table 1, indicate a clear advantage of two contact-based models (EC and PC) over two simpler, distance-based models (EX and SC). From the intuitive point of view, this may be attributed to the fact that contact-based models emphasize base-pairing interactions and are also sensitive to the destabilization of a double helix due to flipped out bases. Among contact-based models, the EC model tends to provide more consistent results than PC model: it performs as best or second-best in almost all categories (unlike PC model which gave the worst results among all models in structure preservation test for DNA) and achieves optimal results for the same value of control parameter for RNA and DNA structures. Among classic, distance-based ENMs, the EC model overperforms SC model, indicating a disadvantage of an unphysical, sharp cutoff. It is worth stressing that once generated, all ENMs considered here require the same computational effort in order to produce a set of normal modes for a given structure. In our view it advocates the use contact-based models, in particular the EC model, whose construction is just slightly more difficult in comparison to standard distance-based models.

**Table 1. Summary of Optimal Control Parameter Values (eqs 2−5), Corresponding Force Constants (in kcal/mol/Å$^2$), and Model Ranks, $N$ (a higher score indicates better performance; see text for details) for RNA and DNA ENMs**

| model | parameter | RNA | DNA | $k_{0,\ \mathrm{RNA}}$ | $k_{0,\ \mathrm{DNA}}$ | $N_{\mathrm{RNA}}$ | $N_{\mathrm{DNA}}$ |
|---|---|---|---|---|---|---|---|
| EC | $d$ [Å] | 5 | 5 | 1.2 | 0.6 | 3 | 5 |
| PC | $n$ | 7 | 4 | $4.7 \times 10^4$ | 16 | 3 | 2 |
| EX | $R^c$ [Å] | 7 | 8 | 150 | 20 | −2 | −2 |
| SC | $R^c$ [Å] | 16 | 16 | 21 | 6 | −4 | −5 |

The apparent importance of base pairing interactions suggests potential benefits of using more than a single bead per nucleotide in order to explicitly build base-centered nodes into the model. Although increasing the resolution of ENMs proved non-beneficial in the case of proteins,[26] systematic evaluation of finer RNA and DNA models may be worth pursuing due to distinct architecture of nucleic acids. One should bear in mind, however, that in general, increasing the resolution of potential energy landscape may reduce the amplitude of structurally correct deformations accessible under the limit of harmonic approximation.

In spite of seemingly clear ranking the differences in absolute performance of the four considered ENMs are rather small (however, all perform undoubtedly better than random ENM). It suggests that all elastic models, in which the net force acting on interaction centers is governed by the distribution of surrounding mass, tend to provide similar baseline efficiency, leaving very small room for improvements due to variations in the range dependence of spring constants. Indeed, apart from the four presented spring formalisms, we tested several others, including some with different scaling of force constants for "bonded" and "nonbonded" bead−bead contacts or with force constants following distance-dependent amplitudes of pseudoelastic force coefficients resulting from VSA, only to learn that their performance is not superior enough to justify an increased level of sophistication. Despite the fact that the performance of ENMs seems to be intrinsically limited to only moderate overlap with experimental changes, we note that their simplicity make them still a very valuable tool to investigate general flexibility and dynamic properties of macromolecules. One of the areas in which a coarse-grained nature of ENMs perfectly matches reduced structure representation is macromolecular docking; here a simplified level of description is essential for the efficient evaluation of countless putative complex geometries. A combination of coarse-grained interaction potentials with ENMs of structural flexibility in the context of protein-nucleic acids docking will be the subject of our future research.


dx.doi.org/10.1021/ct400814n | J. Chem. Theory Comput. 2013, 9, 5460−5470

## ■ ASSOCIATED CONTENT

### Supporting Information

contains the lists of all PDB structures used for the evaluation of ENMs, the description of MD simulations protocols, and the illustration of vibrational frequencies resulting from elastic force constants fitting based on direct overlap with Hessian matrices. This information is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: piotr.setny@cent.uw.edu.pl

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors would like to thank Mahmut Kara and Giuseppe La Rosa for providing molecular dynamics trajectories of RNA and DNA fragments. P.S and M.Z. are supported by a grant from Deutsche Forschungsgemeinschaft: DFG ZA153/19-2.

## ■ REFERENCES

(1) Hubbard, R. E.; Neidle, S.; Campbell, S.; Clore, M.; Lilley, D. M. J.; Dyda, F.; van Duyne, G.; Stella, S.; Heiss, J.; Ferre-D'Amare, A. R.; Li, H.; Horvath, M. P.; Horton, N. C.; Hickman, A. B.; Jen-Jacobsen, L.; Marmorstein, R.; Johnson, R. C.; Hong, M.; Berger, J. M.; Dong, K.; Jacobson, L. A.; Stampfl, S.; Rajkowitsch, L.; Semrad, K.; Berman, H.; Lawson, C. L.; Schroeder, R. In *Protein-Nucleic Acid Interactions: Structural Biology*; Rice, P. A., Correll, C. C., Eds.; RSC Biomolecular Sciences; The Royal Society of Chemistry: London, 2008; Vol. *0*; pp P001−P397.
(2) Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. *Annu. Rev. Biochem.* **2010**, *79*, 233−269.
(3) Leontis, N. B.; Lescoute, A.; Westhof, E. *Curr. Opin. Struct. Biol.* **2006**, *16*, 279−287.
(4) van Dijk, M.; van Dijk, A. D. J.; Hsu, V.; Boelens, R.; Bonvin, A. M. J. J. *Nucleic Acids Res.* **2006**, *34*, 3317−3325.
(5) van Dijk, M.; Bonvin, A. M. J. J. *Nucleic Acids Res.* **2010**, *38*, 5634−5647.
(6) Setny, P.; Zacharias, M. *Nucleic Acids Res.* **2011**, *39*, 9118−9129.
(7) Setny, P.; Bahadur, R. P.; Zacharias, M. *BMC Bioinformatics* **2012**, *13*, 228.
(8) Zacharias, M. *Proteins* **2004**, *54*, 759−767.
(9) Garcia, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696.
(10) Amadei, A.; Linssen, A.; Berendsen, H. J. *Proteins: Struct., Funct., Bioinf.* **1993**, *17*, 412−425.
(11) Brooks, B. R.; Janei, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522−1542.
(12) Dykeman, E. C.; Sankey, O. F. *J. Phys.: Condens. Matter* **2010**, *22*, 423202.
(13) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905.
(14) Tama, F.; Sanejouand, Y.-H. *Protein Eng.* **2001**, *14*, 1−6.
(15) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505−515.
(16) Yang, L.; Song, G.; Jernigan, R. L. *Biophys. J.* **2007**, *93*, 920−929.
(17) Ahmed, A.; Villinger, S.; Gohlke, H. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 3341−3352.
(18) Hinsen, K. *Proteins* **1998**, *33*, 417−429.
(19) Bahar, I.; Rader, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586−592.
(20) Bahar, I.; Jernigan, R. L. *J. Mol. Biol.* **1998**, *281*, 871−884.
(21) Viduna, D.; Hinsen, K.; Kneller, G. *Phys. Rev. E* **2000**, *62*, 3986.
(22) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. *J. Struct. Biol.* **2004**, *147*, 302−314.
(23) Van Wynsberghe, A. W.; Cui, Q. *Biophys. J.* **2005**, *89*, 2939−2949.
(24) Fulle, S.; Gohlke, H. *Biophys. J.* **2008**, *94*, 4202−4219.
(25) Kurkcuoglu, O.; Kurkcuoglu, Z.; Doruker, P.; Jernigan, R. L. *Proteins: Struct., Funct., Bioinf.* **2009**, *75*, 837−845.
(26) Leioatts, N.; Romo, T. D.; Grossfield, A. *J. Chem. Theory Comput.* **2012**, *8*, 2424−2434.
(27) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. *Bioinformatics* **2007**, *23*, 2625−2627.
(28) Case, D. A., et al., *AMBER 12*; University of California: San Francisco, CA, 2012,
(29) Lu, X.-J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108−5121.
(30) Zheng, W.; Brooks, B. R. *Biophys. J.* **2005**, *89*, 167−178.
(31) Woodcock, H. L.; Zheng, W.; Ghysels, A.; Shao, Y.; Kong, J.; Brooks, B. R. *J. Chem. Phys.* **2008**, *129*, 214109.
(32) Hess, B. *Phys. Rev. E* **2002**, *65*, 031910.
(33) Ghysels, A.; Miller, B. T.; Pickard, F. C.; Brooks, B. R. *J. Comput. Chem.* **2012**, *33*, 2250−2275.
(34) Hinsen, K.; Petrescu, A.-J.; Dellerue, S.; Bellissent-Funel, M.-C.; Kneller, G. R. *Chem. Phys.* **2000**, *261*, 25−37.
(35) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.
(36) van Dijk, M.; Bonvin, A. M. J. J. *Nucleic Acids Res.* **2008**, *36*, e88.
(37) Orellana, L.; Rueda, M.; Ferrer-Costa, C.; Lopez-Blanco, J. R.; Chacon, P.; Orozco, M. *J. Chem. Theory Comput.* **2010**, *6*, 2910−2923.
(38) Lu, M.; Ma, J. *Biophys. J.* **2005**, *89*, 2395−2401.
(39) Erman, B. *Biophys. J.* **2006**, *91*, 3589−3599.
(40) Perez, A.; Noy, A.; Lankas, F.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2004**, *32*, 6144−6151.
(41) Noy, A.; Perez, A.; Lankas, F.; Javier Luque, F.; Orozco, M. *J. Mol. Biol.* **2004**, *343*, 627−638.