Empirical corrections to the Amber RNA force field with Target Metadynamics

Alejandro Gil-Ley, Sandro Bottaro, and Giovanni Bussi*

Scuola Internazionale Superiore di Studi Avanzati (SISSA), via Bonomea 265, 34136, Trieste, Italy
E-mail: bussi@sissa.it

1 Simulation Protocols

1.1 Comparison with experimental data

The vicinal nuclear spin-spin $^3J$ scalar couplings constants were calculated according to the conformation of the related torsion angles using the Karplus relationship in the form $J(\phi) = A \cos^2(\phi + \varphi) + B \cos(\phi + \varphi) + C$.\cite{1,2} Several sets of coefficients are available for each specific observable/torsion equation (see ref\cite{3}) and there has been no clear consensus on which of them is to be preferred. We took into account the analysis made in refs.\cite{4–6} to select the most precise sets of parameters. For $^3J_{H_4'\,H_5'}$ and $^3J_{H_4'\,H_5''}$ we derived a simplified expression for the generalized Karplus equation in the form $J(\phi) = A \cos^2(\phi) + B \cos(\phi) + \tilde{B} \sin(\phi) \cos(\phi) + C$. The parameters used in this study are listed in Table 1.

The $^3J$ scalar couplings from the simulations were calculated as the ensemble average over the sampled conformational space, using the following equation

$$\langle J \rangle = \sum_{\theta=-\pi}^{\pi} f(\theta) J(\theta) \delta\theta$$

(1)
Table 1: Karplus parameters for the dihedral angles considered in this study. $\chi'$ indicates the H1'-C1'-N1/9-C6/8 torsion along with a phase shift of 60°, which in the special case of base planar at N1/9 is equal to $\chi$. Actually, the relations of $^{3}J_{C-H}$ with the $\chi$ angle have been shown to depend non-trivially on the sugar pucker and on the nonplanarity of nucleobases.

| Coupling | Angle | A     | B     | C     | $\varphi$ | Ref |
|----------|-------|-------|-------|-------|-----------|-----|
| $^{3}J_{H_1'H_2'}$, $^{3}J_{H_2'H_3'}$, $^{3}J_{H_3'H_4'}$ | $\nu_{1,2,3}$ | 9.67  | -2.03 | 0     | 0         | 6.1 |
| $^{3}J_{H_1'H_2'}$ | $\gamma$ | 8.31  | -0.99 | 0.27  | 1.37      | 60° |
| $^{3}J_{H_2'H_3'}$ | $\gamma$ | 8.31  | -0.99 | -4.72 | 1.37      | 0   |
| $^{3}J_{H_3'H_4'}$ | $\beta$ | 18.1  | -4.8  | 0     | 0         | -120°|
| $^{3}J_{H_4'H_5'}$ | $\beta$ | 18.1  | -4.8  | 0     | 0         | 120° |
| $^{3}J_{H_5'H_6'}$ | $\epsilon$ | 15.3  | -6.1  | 0     | 1.6       | 120° |
| $^{3}J_{C_2'H_6'}$ | $\epsilon$ | 6.9   | -3.4  | 0     | 0.7       | -120°|
| $^{3}J_{C_3'H_5'}$ | $\beta/\epsilon$ | 6.9   | -3.4  | 0     | 0.7       | 0   |
| $^{3}J_{C_3'H_5''}$ | $\chi'$ | 4.8   | 0.7   | 0     | 0.3       | -70.4°|
| $^{3}J_{C_4'H_5'}$ | $\chi'$ | 4.2   | -0.5  | 0     | 0.3       | -68.9°|

where $J(\theta)$ represent the Karplus relation between the vicinal coupling and the dihedral angle and $f(\theta)$ is the probability density of the dihedral angle bin. To calculate the torsion angle histograms we employed wrapped Gaussian kernels with a bandwidth of 0.017 rad. Histograms were calculated with PLUMED 2.1.

The overall agreement between the NMR data and the average values calculated in this study was measured using the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (\langle J_i \rangle_{calc} - J_{i,exp})^2}$$

(2)

The consistency of the error measurements was analyzed by blocking the trajectory in 4 blocks of equal length and calculating the standard deviation of the different error estimations.

2D NOESY experiments of different tetranucleotide sequences have provided rigorous benchmarks for force-fields modifications. NOE distances were calculated by averaging pairwise proton-proton distances over all the structures within the ensemble. The deviation of MD distances from experimental NOE derived distances is calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (\langle J_i \rangle_{calc} - J_{i,exp})^2}$$

(2)
\[ RMSE = \sqrt{\frac{1}{N-1} \sum_{i} \left( \frac{1}{N} \sum_{j=1}^{N} (\frac{1}{r_{ij}})^{\frac{1}{3}} - d_{\text{exp}}^{i} \right)^{2}} \]  

A very important indicator of the ensemble agreement with the experiment is the number of proton-proton contacts with an MD averaged distance of \( \leq 5 \text{ Å} \) which are not visible in the NOESY spectra.\(^7\)

1.2 Thermodynamics

To calculate the free-energy of stacking we used the definition similar to the one of ref\(^7\) to define the stacked and unstacked states. In particular, we calculate the distance between the center of mass of the nucleobases using only the heteroatoms (with a cutoff of 5 Å), the angle between the vectors normal to the planes of the bases (from 0° to 45° and from 135° to 180°) to separate the parallel to the T-shaped complexes, and the angle between the distance vector between the bases and the 5’-nucleobase normal vector (< 50°). This definition is very similar to the one used on ref\(^7\).

1.3 Mutual Information and Jensen-Shannon divergence

The correlation between the dihedral angles in the tetranucleotide T-REMD simulations was estimated with the Mutual Information (MI).\(^17\) We used the driver command of PLUMED 2.1\(^13\) to calculate the MI as an average along the trajectory.

\[ MI_{xy} = \left\langle \log \left[ \frac{p(x,y)}{p(x)p(y)} \right] \right\rangle \]  

\( \text{(4)} \)

The difference between the probability distributions from the A-form and Non-A-form sub-ensembles was measured using the Jensen–Shannon divergence (JS).\(^18,19\) The JS is zero for identical distributions and reaches it maximum (\( \log 2 \)) for non-overlapping ones. The probability distributions used to estimated JS are shown in Figs. S.4, S.5 and S.6.

\[ JS_{AB} = \frac{1}{2} \left\langle \log \left[ \frac{2p_{A}(x,y)}{p_{A}(x,y) + p_{B}(x,y)} \right] \right\rangle_{A} + \frac{1}{2} \left\langle \log \left[ \frac{2p_{B}(x,y)}{p_{A}(x,y) + p_{B}(x,y)} \right] \right\rangle_{B} \]  

\( \text{(5)} \)
2 Selection of the target collective variables

The prevalence of compact intercalated and inverted conformations in the ensembles of RNA tetranucleotides generated with Amber force fields is a known problem (see Fig. S.2 for a representation of typical structures). This can be due to an over-stabilization of stacking interactions, poor water models, and/or incorrect dihedral parameters. Changing the non-bonded interactions in a force field to improve stacking is a difficult task, as the classification of stacked (closed) and non-stacked (open) structures in a molecular dynamic simulation is largely arbitrary and slight changes can lead to very different values of the open-closed population ratio. On the other hand, dihedrals terms are more flexible and small corrections in the free-energy profiles of a minimal number of angles can have huge impact on the whole nucleic acids ensemble. So, we decided to correct the free energy landscape of an essential group of dihedral angles in the RNA, in order to improve the state-of-the-art Amber force field agreement with solution NMR data.

For this analysis we used the T-REMD simulations of AAAA, GACC, and CCCC tetranucleotides performed on ref using the Amber14 force field. We divided each of the Amber14 ensembles into two groups, in order to identify the structural features that differentiate the structures compatible with the NMR data, A-form-like conformation, from the non-compatible compact structures that overpopulate the ensembles. The A-form sub-ensemble was defined as the set of conformations with a distance-RMSD < 2.5 Å from the canonical A-form, while the Non-A-form group comprises the rest of the frames. The ratio between the population of the Non-A-form sub-ensemble over the A-form one is different for each tetranucleotide: \(~5.5\) for AAAA, \(~1.4\) for GACC and \(~21.7\) for CCCC.

Differences among the collective variable (CV) distributions of the sub-ensembles were measured using the Jensen-Shannon (JS) divergence between the two-dimensional probability distributions of several CVs (dihedrals ($\alpha \to \chi$), puckering coordinates ($Z_x$) and the nucleobase-nucleobase coordination number ($S$): $\alpha\beta$, $\beta\gamma$, $\gamma Z_x$, $Z_x \chi$, $S \chi$, $Z_x \epsilon$, $\epsilon \zeta$, $\zeta \alpha$). Coordination numbers were estimated using a switching function with form \(\frac{1}{1+(r/r_0)^6}\) with \(r_0 = 0.3\)nm. In Fig. S.3 it can be appreciated the JS divergences for each pair of CVs,
while the probability distribution maps employed in the JS calculation are shown in Fig S.4, S.5 and S.6. The pair of CVs with the highest JS values are the ones containing $\chi$, $\zeta$ and $\alpha$. It should be noticed that JS values in AAAA for the $Z_x \chi$ and $S \chi$ pairs are in general higher than the ones corresponding to the GACC and CCCC. Analysis of the probability distributions shows that in the case of AAAA the $\chi$ angle in the A-form sub-ensemble favors the high-anti and syn conformations instead of the canonical A-form all anti rotamer, while in the Non-A-form group the minima are shifted to the anti state.

For the tetranucleotides containing pyrimidines the $\chi$ angles mainly populates the anti state in both sub-ensembles. This result could suggest some problematic behavior of the $\chi$ angle in adenosine that should be further investigated [see ref20 for a discussion of problematic behavior of $\chi$ angle in RNA tetraloops]. A consistent trend among all tetranucleotide is related to the high JS values for $\alpha\beta$, $\epsilon\zeta$ and $\zeta\alpha$ dihedral pairs. Looking into the probability distributions it is clear that the high JS divergence values are due to a shift of the $\zeta\alpha$ minimum from the $\zeta(-g)/\alpha(-g)$ (the one corresponding to the right-handed helix) in the A-form sub ensemble, to the $\zeta(+g)/\alpha(+g)$ conformation in the Non-A-form. Each $\zeta\alpha$ minimum seems to be characteristic of each sub-ensemble independently of the RNA sequence, while $\chi$ and stacking are very system dependent. Taking these results into account we hypothesize that changing the stability of the gauche minima in $\zeta/\alpha$ can improve the experimental agreement of the force field, as most conformations compatible with the canonical A-form extended structure will be favored, despite the properties of the sugar-base domain not been changed.

The $\zeta$ and $\alpha$ torsion angles are highly correlated between each other and with the $\epsilon$ and $\beta$ angles respectively, as appreciated in Fig S.3. Therefore, we assume that any modification on the phosphodiester backbone should include also the adjacent torsions.

The probability distributions of the RNA backbone angles obtained from the PDB can be a good reference to correct the Amber14 force field, as long as those distributions are compatible with the solution RNA ensemble at room temperature. In order to analyze the suitability of the PDB distributions, we used solution NMR data of RNA dinucleosides as a reference. RNA dinucleoside monophosphates can be considered as the smallest
structural unit of the RNA that includes all the major conformational degrees of freedom. Thanks to their small size, converged ensembles are easily generated using enhanced sampling simulations. Moreover, taking fragments of dinucleotides from the RNA X-ray structures, instead of tetranucleotides, improves considerably the statistics. In Fig S.7 the agreement between experimental and calculated $^3J$ scalar couplings for the dinucleosides is shown. For the X-ray ensemble in general the agreement with the scalar couplings of the backbone angles ($\epsilon$ and $\beta$) is better than that of angles of the ribose-nucleobase region ($\chi$, $Z_x$ and $\gamma$). The disagreement in the last region is expected considering the X-ray ensemble is biased to the anti and C$_{3\text{-}endo}$ states, which predominate on the double helical structures. Compared to the force field performance, the $^3J$ RMSD of the PDB fragments is at least 0.5 Hz lower for the backbone angles. All the calculated and experimental $^3J$ scalar couplings used are presented in Table 2.

We decided then to enforce the X-ray distributions of $\alpha$, $\beta$, $\epsilon$, and $\varsigma$ dihedral angles in the Amber14 force field, using concurrent Target Metadynamics simulations.

### 3 Results

#### List of Figures

- **S.1** T-MetaD+RECT input file. .................................................. 8
- **S.2** Representative clusters of the tetranucleotide Amber14 ensemble (AAAA). 9
- **S.3** Jensen-Shannon divergence and Mutual Information. .................... 10
- **S.4** Probability distributions of A-form and Non-A-form sub-ensembles (AAAA). 11
- **S.5** Probability distributions of A-form and Non-A-form sub-ensembles (CCCC). 12
- **S.6** Probability distributions of A-form and Non-A-form sub-ensembles (GACC). 13
- **S.7** Free-energy profiles. ............................................................... 14
- **S.8** Probability distributions of backbone angles (AA). ........................ 15
- **S.9** Probability distributions of backbone angles (CC). ........................ 16
- **S.10** Probability distributions of backbone angles (AC). ........................ 17
S.11 Probability distributions of backbone angles (CA) . . . . . . . . . . . . . . 18
S.12 Free-energy profiles of non-corrected degrees of freedom. . . . . . . . . . 19
S.13 RMSE between experimental and calculated $^3J$ scalar couplings. . . . . 20
S.14 Free Energy of stacking. . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
S.15 Predicted versus experimental NOE distance . . . . . . . . . . . . . . . . 23
S.16 Empirical RMSD probability distribution. . . . . . . . . . . . . . . . . . 24
S.17 Probability distributions of backbone dihedral angles (AAAA/AA). . . . . 25
S.18 Probability distributions of backbone dihedral angles (AAAA/AA). . . . . 26
S.19 Probability distributions of backbone dihedral angles (AAAA/AA). . . . . 27
S.20 Probability distributions of backbone dihedral angles (CCCC/CC). . . . . 28
S.21 Probability distributions of backbone dihedral angles (CCCC/CC). . . . . 29
S.22 Probability distributions of backbone dihedral angles (CCCC/CC). . . . . 30
S.23 Probability distributions of backbone dihedral angles (GACC/AA/AC/CC). 31
S.24 Probability distributions of backbone dihedral angles (GACC/AA/AC/CC). 32
S.25 Probability distributions of backbone dihedral angles (GACC/AA/AC/CC). 33
S.26 RMSD distribution as a function of the bias potential strength. . . . . . 34

List of Tables

1 Karplus relation parameters . . . . . . . . . . . . . . . . . . . . . . . . . 22
2 Scalar couplings. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
Figure S.1: Sample input file for a T-MetaD and RECT simulation of the AC dinucleoside. The section “Concurrent Metadynamics” in the input file should include all the CVs defined before, but it is abbreviated in this example to only 3 of them.
Figure S.2: Representative clusters of the tetranucleotide Amber14 ensemble of AAAA. The clusters were calculated with the gromos algorithm implemented in the g_cluster tool of Gromacs. Representative structures for the CCCC and GACC can be appreciated in ref., for a highly sampled Amber14 ensemble.
Figure S.3: Jensen-Shannon (JS) divergence and Mutual Information (MI) calculated from the joint probability distributions of the CV1CV2 pairs indicated in the labels. The distributions were estimated from the tetranucleotides Amber14 ensembles taken from ref[24]. The Amber14 ensemble was divided into two groups, containing the A-form-like and Non-A-form structures respectively. The Jensen-Shannon divergence measures the difference between the bidimensional CV1CV2 probability distributions from the A-form-like and Non-A-form structures. The Mutual Information was calculated for the full Amber14 ensemble. MI indicates the correlation between the CV1 with respect to the CV2. The shaded areas represent the JS and MI values obtained for a random generated set of data of the same size of the Amber14 ensemble. Those values differ from zero due to the finite size of the datasets. The significance of the calculated JS and MI values is proportional to their distance from the shaded area.
Figure S.4: Probability distributions of dihedral angles (backbone, puckering and glycosidic angle) and coordination number of the nucleobases center-of-mass distance. These CVs values corresponding to the RNA canonical A-form are marked with a gray dashed line. The probability distributions were calculated from the AAAA T-REMD simulations. The distributions marked as “A-form” includes the structures with a distance RMSD \( \leq 2.5 \) Å to the canonical A-form conformation in the Amber14 ensemble, while the “Non A-form” group contains the rest, mostly compact and highly stacked structures.
Figure S.5: Same as Fig. S.4 but for the CCCC Amber14 ensemble.
Figure S.6: Same as Fig. S.4 but for the GACC Amber14 ensemble.
Figure S.7: Free-energy profiles of backbone dihedral angles for all the dinucleosides monophosphates studied here, from the X-ray ensemble (PDB) and the RECT simulations with the standard force-field (Amber14) and the correcting potential (Amber\_pdb).
Figure S.8: Probability distributions of the backbone dihedral angles of AA dinucleoside monophosphate, from the X-ray ensemble (PDB) and the RECT simulations with the standard force-field (Amber14) and the correcting potential (Amber\_pdb).
Figure S.9: Probability distributions of the backbone dihedral angles of CC dinucleoside monophosphate, from the X-ray ensemble (PDB) and the RECT simulations with the standard force-field (Amber14) and the correcting potential (Amber_{pdb}).
Figure S.10: Probability distributions of the backbone dihedral angles of AC dinucleoside monophosphate, from the X-ray ensemble (PDB) and the RECT simulations with the standard force-field (Amber14) and the correcting potential (Amber_{pdb}).
Figure S.11: Probability distributions of the backbone dihedral angles of CA dinucleoside monophosphate, from the X-ray ensemble (PDB) and the RECT simulations with the standard force-field (Amber14) and the correcting potential (Amber\textsubscript{pdb}).
Figure S.12: Free-energy profiles of non-corrected degrees of freedom ($\chi$, $\gamma$ and puckering $Z_x$) from the RECT simulations of the standard force-field (Amber14) and the correcting potential (Amber$_{pdb}$).
Figure S.13: RMSE between experimental and calculated $^3J$ scalar couplings for 2 different subset of dihedral angles: Nucs.) containing the nucleoside-unit angles ($\chi$, $\gamma$ and $\nu_3$) and Back.) including the angles from the monophosphate backbone ($\epsilon$, $\zeta$, $\alpha$ and $\beta$).
| seq. | source | unit | \( \nu_{3-1}/\nu_{3-2} \) | \( \gamma_{1}/\gamma_{2} \) | \( \beta_{1}/\beta_{2} \) | \( \epsilon_{1}/\beta_{2} \) | \( \chi_{1}/\chi_{2} \) |
|------|--------|------|-------------------|-------------------|-------------------|-------------------|-------------------|
|      |        |      |                   |                   |                   |                   |                   |
| 1 1  | NMR*   |      | 7.3/7.0           | 2.5/2.5           | 3.8/3.3           | 8.9/8.3           | 3.1               | 6.0               | 1.4               | 4.5               |
| 2 1  |        |      | 7.2/6.6           | 2.4/2.2           | 2.4/2.6           | -                 | 4.3/4.0           | 3.2/3.6           | -                 | 9.5               | 1.4               | 4.6               |
| 1 2  | Amber4 |      | 9.4               | 3.8               | 2.9               | 6.8               | -                 | -                 | 2.3               | 7.6               | 1.7               | 2.9               |
| 2 2  |        |      | 7.1               | 3.3               | 1.3               | -                 | 3.4               | 1.9               | -                 | 10.5              | 1.8               | 3.3               |
| 1 3  | PDB    |      | 6.8               | 4.4               | 2.4               | 7.6               | -                 | -                 | 1.2               | 8.3               | 0.8               | 1.9               |
| 2 3  |        |      | 7.1               | 4.5               | 2.4               | -                 | 3.3               | 2.6               | -                 | 10.3              | 0.9               | 1.9               |
| 1 4  | Amberpdb |     | 10.4              | 3.8               | 2.9               | 7.6               | -                 | -                 | 1.4               | 8.2               | 1.3               | 2.5               |
| 2 4  |        |      | 6.1               | 3.4               | 1.3               | -                 | 3.6               | 2.2               | -                 | 10.4              | 1.7               | 3.3               |
|      |        |      |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |
| 1 5  | NMR†   |      | ~5.0/5.5          | 2.5/2.5           | 3.6/3.5           | ~9.0/7.6          | -                 | -                 | 3.7               | 5.3               | 1.9               | 4.2               |
| 2 5  |        |      | 5.5/5.8           | ~2.8              | ~3.8/3.7          | ~3.0              | ~3.8/3.3          | -                 | 2.5               | 3.1               |                   |                   |
| 1 6  | Amber4 |      | 7.6               | 3.4               | 1.5               | 5.6               | -                 | -                 | 3.1               | 7.3               | 4.3               | 3.5               |
| 2 6  |        |      | 6.2               | 3.6               | 1.6               | -                 | 3.1               | 2.5               | -                 | 10.4              | 1.7               | 2.7               |
| 1 7  | PDB    |      | 6.8               | 4.8               | 3.3               | 8.4               | -                 | -                 | 2.9               | 6.3               | 1.2               | 2.7               |
| 2 7  |        |      | 7.1               | 4.5               | 2.8               | -                 | 4.9               | 3.5               | -                 | 9.2               | 1.3               | 2.8               |
| 1 8  | Amberpdb |     | 7.2              | 3.4               | 1.7               | 8.1               | -                 | -                 | 3.1               | 6.3               | 3.8               | 3.1               |
| 2 8  |        |      | 5.4               | 3.9               | 1.9               | -                 | 4.2               | 2.9               | -                 | 9.8               | 1.6               | 2.8               |
|      |        |      |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |
| 1 9  | NMR†   |      | 6.1/6.1           | 2.4/3.3           | 3.5/2.2           | 8.7/8.7           | -                 | -                 | 3.3               | 4.6               | 2.1               | 2.8               |
| 2 9  |        |      | ~7.1              | ~1.7/2.1          | ~2.0/2.6          | ~4.0/3.6          | 3.4/3.6           | -                 | 9.5               | 1.4               | 4.5               |                   |
| 1 10 | Amber4 |      | 8.4               | 2.9               | 1.9               | 6.1               | -                 | -                 | 2.5               | 7.7               | 3.4               | 2.8               |
| 2 10 |        |      | 5.9               | 3.4               | 1.4               | -                 | 3.1               | 2.4               | -                 | 10.5              | 1.6               | 3.1               |
| 1 11 | PDB    |      | 7.7               | 4.5               | 2.7               | 8.5               | -                 | -                 | 2.4               | 6.8               | 1.0               | 2.4               |
| 2 11 |        |      | 8.2               | 4.5               | 2.6               | -                 | 4.5               | 2.8               | -                 | 9.7               | 1.0               | 2.1               |
| 1 12 | Amberpdb |     | 8.0              | 3.3               | 2.3               | 7.9               | -                 | -                 | 2.7               | 6.7               | 2.8               | 2.6               |
| 2 12 |        |      | 5.6               | 3.6               | 1.5               | -                 | 3.8               | 2.4               | -                 | 10.2              | 1.6               | 3.2               |

Table 2: Scalar couplings for the monophosphate dinucleosides. The coupling constants from the torsion angles of the backbone are highlighted in red. Experimental values were taken from refs (\textsuperscript{5}) \textsuperscript{21}23. \textsuperscript{*}(Experiments temperature: 320/293 K). \textsuperscript{†}(280/293 K).
Figure S.14: Free Energy of stacking. Experimental values were taken from refs \cite{28,29}. The stacking thermodynamics is only slightly affected by the correcting potentials.
Figure S.15: Predicted versus experimental NOE distance for all ensembles and for all 5 systems are shown below. Bars on the y-axis show experimentally determined minimum and maximum range, while error bars on the predicted values represent statistical errors and were calculated with a blocking procedure. Intra-nucleotide and inter-nucleotide proton-proton distances are shown in blue and green, respectively. Calculations were performed using the software tool baRNAba.\textsuperscript{[16]}
Figure S.16: Empirical probability distribution of heavy atom RMSD from the canonical A-form as computed for the reference replica trajectory. Distributions are shown for the REMD simulations with Amber14 (black) and Amber_{pdb} (red). The total probability is shown in solid line and the above and below limits determined by the blocking error are shaded. It can be appreciated that the correcting potentials increase the population of extended structures (RMSD $\sim$ 1-2 Å) for the CCCC and GACC tetranucleotides, while for AAAA the Amber_{pdb} ensemble is farther from the canonical A-form.
Figure S.17: Probability distributions of the backbone dihedral angles of AAAA tetranucleotide, in the region between residue 1 and 2. First are shown the dihedral distributions from the RECT simulations with the standard force-field (Amber14) and in second the ones performed with the correcting potential (Amber$_{pdb}$). The dihedral distributions of AA dinucleoside taken from the PDB are presented in the last row.
Figure S.18: Probability distributions of the backbone dihedral angles of AAAA tetranucleotide, in the region between residue 2 and 3. First are shown the dihedral distributions from the RECT simulations with the standard force-field (Amber14) and in second the ones performed with the correcting potential (Amber\_pdb). The dihedral distributions of AA dinucleoside taken from the PDB are presented in the last row.
Figure S.19: Probability distributions of the backbone dihedral angles of AAAA tetranucleotide, in the region between residue 3 and 4. First are shown the dihedral distributions from the RECT simulations with the standard force-field (Amber14) and in second the ones performed with the correcting potential (Amber_{pdb}). The dihedral distributions of AA dinucleoside taken from the PDB are presented in the last row.
Figure S.20: Probability distributions of the backbone dihedral angles of CCCC tetranucleotide, in the region between residue 1 and 2. First are shown the dihedral distributions from the RECT simulations with the standard force-field (Amber14) and in second the ones performed with the correcting potential (Amberpdb). The dihedral distributions of CC dinucleoside taken from the PDB are presented in the last row.
Figure S.21: Probability distributions of the backbone dihedral angles of CCCC tetranucleotide, in the region between residue 2 and 3. First are shown the dihedral distributions from the RECT simulations with the standard force-field (Amber14) and in second the ones performed with the correcting potential (Amber_{pdb}). The dihedral distributions of CC dinucleoside taken from the PDB are presented in the last row.
Figure S.22: Probability distributions of the backbone dihedral angles of CCCC tetranucleotide, in the region between residue 3 and 4. First are shown the dihedral distributions from the RECT simulations with the standard force-field (Amber14) and in second the ones performed with the correcting potential (Amber_{pdb}). The dihedral distributions of CC dinucleoside taken from the PDB are presented in the last row.
Figure S.23: Probability distributions of the backbone dihedral angles of GACC tetranucleotide, in the region between residue 1 and 2. First are shown the dihedral distributions from the RECT simulations with the standard force-field (Amber14) and in second the ones performed with the correcting potential (Amber_{pdb}). The dihedral distributions of AA, AC and CC dinucleosides, taken from the PDB, are presented in the last row.
Figure S.24: Probability distributions of the backbone dihedral angles of GACC tetranucleotide, in the region between residue 2 and 3. First are shown the dihedral distributions from the RECT simulations with the standard force-field (Amber14) and in second the ones performed with the correcting potential (Amber_pdb). The dihedral distributions of AA, AC and CC dinucleosides, taken from the PDB, are presented in the last row.
Figure S.25: Probability distributions of the backbone dihedral angles of GACC tetranucleotide, in the region between residue 3 and 4. First are shown the dihedral distributions from the RECT simulations with the standard force-field (Amber14) and in second the ones performed with the correcting potential (Amber_{pdb}). The dihedral distributions of AA, AC and CC dinucleosides, taken from the PDB, are presented in the last row.
Figure S.26: Empirical probability distribution of heavy atom RMSD from the canonical A-form computed for the reweighted Amber14 ensemble as a function of the Gaussian potential height. For all sequences the increase of the penalty potential shifts the distributions closer to the A-form structure.
References

(1) Karplus, M. J. Chem. Phys. 1959, 30, 11–15.
(2) Karplus, M. J. Am. Chem. Soc. 1963, 85, 2870–2871.
(3) Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H. Chembiochem 2003, 4, 936–962.
(4) Sychrovský, V.; Vokáčová, Z.; Šponer, J.; Špacková, N.; Schneider, B. J. Phys. Chem. B 2006, 110, 22894–22902.
(5) Vokáčová, Z.; Bickelhaupt, F. M.; Šponer, J.; Sychrovský, V. J. Phys. Chem. A 2009, 113, 8379–8386.
(6) Vokáčová, Z.; Budesinsky, M.; Rosenberg, I.; Schneider, B.; Šponer, J.; Sychrovský, V. J. Phys. Chem. B 2009, 113, 1182–1191.
(7) Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H. J. Chem. Theory Comput. 2015, 11, 2729–2742.
(8) Haasnoot, C.; de Leeuw, F. A.; Altona, C. Tetrahedron 1980, 36, 2783–2792.
(9) Lee, C.-H.; Sarma, R. H. J. Am. Chem. Soc. 2003, 125, 3649–3658.
(10) Munzarová, M. L.; Sklenár, V. J. Am. Chem. Soc. 2003, 125, 3649–3658.
(11) Vokáčová, Z.; Trantírek, L.; Sychrovský, V. J. Phys. Chem. A 2010, 114, 10202–10208.
(12) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. Comput. Phys. Commun. 2014, 185, 604–613.
(13) Yildirim, I.; Stern, H. A.; Tubbs, J. D.; Kennedy, S. D.; Turner, D. H. J. Phys. Chem. B 2011, 115, 9261–9270.
(14) Tubbs, J. D.; Condon, D. E.; Kennedy, S. D.; Hauser, M.; Bevilacqua, P. C.; Turner, D. H. Biochemistry 2013, 52, 996–1010.
(15) Condon, D. E.; Yildirim, I.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Turner, D. H. J. Phys. Chem. B 2014, 118, 1216–1228.
(16) Cover, T. M.; Thomas, J. A. Elements of Information Theory 1991, 279–335.
(17) Lin, J. Information Theory, IEEE Transactions on 1991, 37, 145–151.
(18) Endres, D. M.; Schindelin, J. E. IEEE Transactions on Information theory 2003.
(19) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. RNA 2015, 21, 1578–1590.
(20) Ezra, F. S.; Lee, C.-H.; Kondo, N. S.; Danyluk, S. S.; Sarma, R. H. Biochemistry 1977, 16, 1977–1987.
(21) Lee, C.-H.; Ezra, F. S.; Kondo, N. S.; Sarma, R. H.; Danyluk, S. S. Biochemistry
1976, 15, 3627–3639.
(30) Frechet, D.; Ehrlich, R.; Remy, P.; Gabarro-Arpa, J. Nucleic Acids Res. 1979, 7, 1981–2001.
(31) Bottaro, S.; Di Palma, F.; Bussi, G. Nucleic Acids Res. 2014, 42, 13306–14.