



NARall: a novel tool for reconstruction of the all-atom structure of nucleic acids from heavily coarse-grained model

Łukasz Golon¹ · Adam K. Sieradzan¹

Received: 19 August 2022 / Accepted: 13 December 2022 / Published online: 22 December 2022
© The Author(s) 2022

Abstract

Nucleic acids are one of the most important cellular components. These molecules have been studied both experimentally and theoretically. As all-atom simulations are still limited to short time scales, coarse-grain modeling allows to study of those molecules on a longer time scale. Nucleic-Acid united RESidue (NARES-2P) is a low-resolution coarse-grained model with two centers of interaction per repeating unit. It has been successfully applied to study DNA self-assembly and telomeric properties. This force field enables the study of nucleic acids Behavior on a long time scale but lacks atomistic details. In this article, we present new software to reconstruct atomistic details from the NARES-2P model. It has been applied to RNA pseudoknot, nucleic acid four-way junction, G-quadruplex and DNA duplex converted to NARES-2P model and back. Moreover, it was applied to DNA structure folded and self-assembled with NARES-2P.

Keywords Nucleic acids · All-atom structure reconstruction · NARES-2P

Introduction

DNA and RNA molecules are essential for life. For example, mice lacking a precise DNA repair mechanism exhibit multi-level abnormalities Wilson and Thompson (1997). The studies of DNA and RNA conformational changes and self-organization are extremely computationally demanding. Even a small RNA loop, such as UUCG poses a significant challenge for computational methods Mráziková et al. (2020). The use of molecular dynamics in all-atom representation has been applied to study non-coding RNA functions (Palermo et al. 2019), RNA and DNA self-assembly (Yoo et al. 2020) and local DNA flexibility (Mazur 2011). However, for the huge systems (over 1,000,000 atoms) time scales are limited to the μ s range Yan et al. (2015), Hang et al. (2015). Therefore, for systems requiring longer time scales, coarse-graining is needed. In coarse-grained force fields, several atoms are merged into one center of interaction enabling a significant speed-up of the conformational sampling due to the smaller number of interaction centers and smoothening of the potential energy surface.

Coarse-grain application to various macromolecules has been described in a recent review Dhamankar and Webb (2021).

CG models of biomolecules are constructed in a wide range of resolutions. One of the most commonly used model is Martini. For nucleic acids, it uses 6-7 interaction sites per nucleotide Uusitalo et al. (2015). PRIMONA, a CG model with 12-13 interaction sites per nucleotide can retain atomistic accuracy, although the simulation speed-up is moderate Gopal et al. (2010). Lower-resolution models include 3SPN (named for having 3 sites per nucleotide) (Sambriski et al. 2009; Hinckley et al. 2013) and oxDNA (with four sites per nucleotide) (Ouldrige et al. 2010), which can handle longer polynucleotide chains. One of the most promising models is the coarse-grained (CG), Nucleic-Acid united RESidue 2-points (NARES-2P) model which has only two centers of interaction per repeating unit (He et al. 2013; Liwo et al. 2014; He et al. 2015). It has been successfully applied to RNA and DNA structure prediction (Sieradzan et al. 2018), DNA association studies (He et al. 2013) and telomere properties predictions (Sieradzan et al. 2017). It can speed up the simulation by several orders of magnitude as in this model several atoms are merged into one center of interaction. NARES-2P is part of the CG force field family UNICORN (Liwo et al. 2021). For a part of UNICORN dedicated

✉ Adam K. Sieradzan
adam.sieradzan@ug.edu.pl

¹ Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland

to proteins (UNRES) 4000-speed up is obtained (Khalili et al. 2005) in comparison with all-atom simulation. However, CG has several drawbacks: Usually, the accuracy of the structure is lower than for an all-atom (AA) model, energy function can be significantly more complicated and temperature-dependent (especially if the model is intended to be transferable), and finally CG model requires back-mapping (reconstruction of AA structure) to regain atomistic details. Multi-scale simulations which use different levels of representation reveal a promising potential (Peter and Kremer 2009). There are known, successful, back-mapping algorithms (Wassenaar et al. 2014; Poblete et al. 2018) for 3(5) interaction centers per repeating unit for SPQR coarse-grained force field (Poblete et al. 2017) or 4 heavy atoms to 1 interaction center MARTINI coarse-grained force field Uusitalo et al. (2015). Another successful, example of back-mapping for a very coarse-grained representation (only 3 centers of interaction per repeating unit) was developed by Shimizu and Takada (2018).

Previously for NARES-2P model, we used the xleap software (Case et al. 2016) to convert the coarse-grained structure to all-atom representation Sieradzian et al. (2017). In that procedure, the position of the center of mass of the sugar ring was assigned as the C1' atom and the position of the nucleoside center was assigned as the C5' atom. This led to significant structure distortion and the recovered structure could only be used for visualization purposes in cartoon representation. In this article, we present new software that restores the AA structure from the NARES-2P model which makes it possible to grasp atomistic details while enabling large times scale simulations. Compared to the back-mapping algorithms referred to above (Wassenaar et al. 2014; Poblete et al. 2018, 2017; Uusitalo et al. 2015) converting the NARES-2P structures to all-atom structures was more challenging since only two interaction sites are present in the NARES-2P model.

Methods

NARES model of the nucleic-acid chain

In the NARES-2P model He et al. (2013); Liwo et al. (2014), a polynucleotide chain is represented by a sequence of deoxyribose (for DNA) or ribose (for RNA) ring centers (S) with united nucleoside (sugar-nucleic base;B) centers attached to them and united phosphate groups (P) located in the midpoint between the consecutive ring centers (Fig. 1). As the Cas in UNRES, the virtual atoms S only serve to define the geometry of a virtual chain, while the B and P centers are the interaction sites.

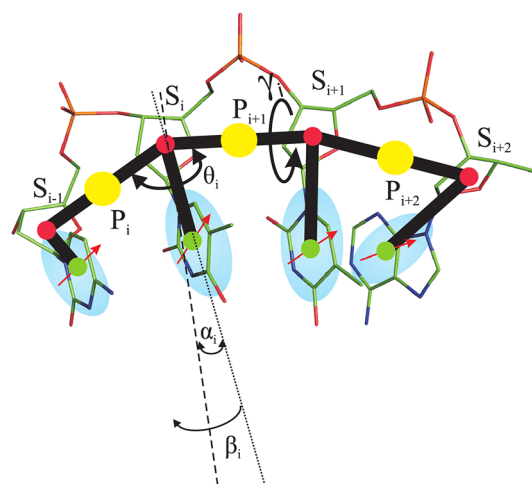


Fig. 1 Comparison between the all-atom representation of the nucleotide chain (sticks) and the coarse-grained NARES-2P model (circles, ellipsoids, and thick black lines). Red circles represent the united sugar groups (S), which serve only as geometric reference points. Yellow circles represent the united phosphate groups (P). Ellipsoids (light blue) represent united nucleosides, with their geometric centers shown as solid green circles. The P's are located halfway between two consecutive sugar atoms. Dipoles are located on the bases to represent their electrostatic interaction. The electrostatic part of the base-base interactions is represented by the mean-field interactions of the base dipoles (red arrows). Virtual-bond angles, θ , and virtual-bond dihedral angles, γ , are used to describe the backbone geometry. The base orientation angles α and torsional angles β , which define the location of a base with respect to the backbone, are also indicated. Adapted from Liwo, A., Baranowski, M., Czaplewski, C. et al. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J. Mol. Model.* 20, 2306 (2014)

Back-mapping algorithm

The algorithm (Fig. 2) for back-mapping starts with reading the coarse-grained coordinates and sequence from a NARES-2P PDB file.

Subsequently, the nucleoside library is read. In the library, the B-DNA and A-RNA nucleosides are used as a reference for the DNA and RNA, respectively. For each nucleoside in the library there is one structure (one rotamer). Nucleosides in the library were constructed with the use of 3DNA (Li et al. 2019) with calf thymus DNA model which corresponds to B_{II} structure. The O5' oxygen atom is not present in the library. As a first step, the nucleosides are reconstructed (Fig. 3). Moving through the chain from 5' to 3'-end nucleoside by nucleoside an all-atom structure is taken from the library. The nucleoside geometries in the library are described in a local coordinate system based on the following four points: the geometric center of the sugar ring (corresponding to S), nucleobase center of mass (corresponding to B), as well as C5' and O3' atoms.

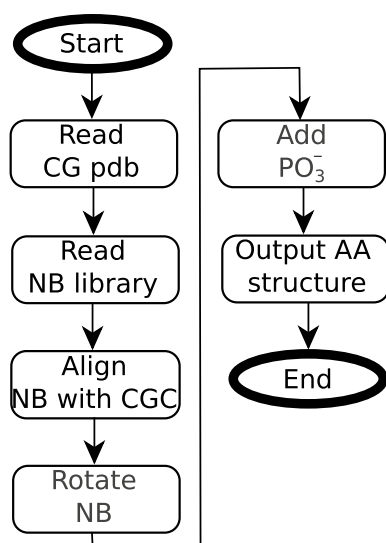


Fig. 2 The backbone reconstruction algorithm block diagram

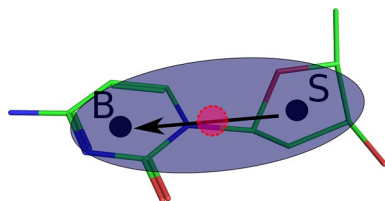


Fig. 3 Superposition of the NARES-2P side-chain (blue spheroid) with the reconstructed all-atom structure from the library. The center of mass of the sugar ring of the all-atom representation and the center of mass of the nucleobase are indicated as black circles and termed S and B, respectively. The center of the NARES-2P side-chain and the NARES-2P sugar-base vector are indicated as a pink circle and black vector, respectively

The local coordinate system is set up so that the X axis is coaxial with the S-B vector, the origin is in the midpoint of the S-B vector, and the Y axis is parallel to the C5'-O3' vector's component normal to the S-B vector.

During reconstruction the nucleosides are placed so that the origin of the local coordinate system is in the midpoint of the CG model's S-B vector, the X axis of the local coordinate system coincides with the CG model's S-B vector, and the Y axis of the local coordinate system is parallel to the CG model's $B_{(i-1)} - B_{(i+1)}$ vector, where i refers to the index number of the residue. In the case of the first and last nucleoside B_i replaces the nonexistent B_{-1} and $B_{(n+1)}$ points (where n is the total number of residues in the oligonucleotide chain) (Fig. 4).

Once the nucleoside is recovered the missing P and O5' atoms are reconstructed in the next step. These atoms are placed on the plane containing the $O3'_i$ and $C5'_{(i+1)}$ atoms

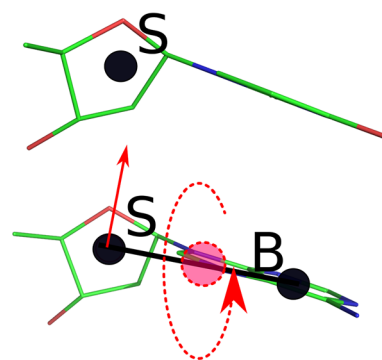


Fig. 4 Schematic representation of sugar-base rotation during back-mapping. The black circle represents the center of mass of the sugar (S) and a nucleobase (B). The pink circle indicates the midpoint of the sugar-base vector. The red arrow is the normal vector to the S-B vector in direction of the previous sugar

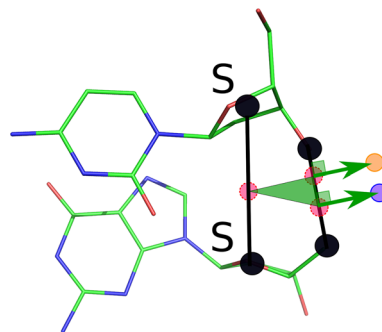


Fig. 5 Schematic representation of backbone phosphorus atom and oxygen atom reconstruction. The pink circles indicate the three points used for P and O5' plane reconstruction. The green triangle determines the plane for placement of the P and O5' atoms and the green arrows are vectors of shifting of P and O5' on the indicated plane from O3' - C5' vector (black line)

reconstructed in the previous step, as well as the midpoint of CG $B_i - B_{(i+1)}$ vector. O5' is placed in 0.735 fraction of all-atom $O3'_i$ and $C5'_{i+1}$ vector and then shifted by 1.32\AA in direction perpendicular to $O3'_i - C5'_{i+1}$ vector and opposite to the NARES-2P B_i to B_{i+1} vector midpoint. P atom is placed in 0.238 fraction of all-atom $O3'_i$ and $C5'_{i+1}$ vector and then shifted by 1.32\AA in direction perpendicular to $O3'_i - C5'_{i+1}$ vector and opposite to the NARES-2P B_i to B_{i+1} vector (Fig. 5). Two remaining oxygen atoms (OP1 and OP2) are remapped with the constraint of the tetrahedral structure of the phosphate group and constant bond length.

The fully reconstructed structure sometimes has clashes and/or inappropriate bond lengths in the sugar-phosphate backbone. Therefore, a short minimization with the use of the AMBER14 ff14SB force field is used with implicit solvent Maier et al. (2015). Firstly, 400 steps of the steepest

descent minimization without any restraints are performed to remove any clashes, bad bond lengths, etc. Secondly, 500 steps of the steepest descent minimization with restraints are performed. The torsional restraints were imposed on the 4 atoms adjacent to: C1', C2', and C3' atoms. The flat bottom restraints with force constant 25 kcal/mol/rad² and the flat bottom with of 20° were imposed. The restraints are based on the initially reconstructed structure. This drives toward the initial NARES-2P structure and allows us to obtain a structure as close to the initial structure without any clashes.

Application of the back-mapping algorithm

We transformed several experimental structures to coarse-grained NARES-2P representation: (a) the pyrimidine motif triple helix in the *Kluyveromyces lactis* telomerase RNA pseudoknot (PDB code: 2M8K) (Cash et al. 2013) (b) the 222-parallel G4 quadruplex (GGTTGGTTGGTTGG) (c) the NMR Structure of a 21 bp DNA duplex preferentially cleaved by Human Topoisomerase II (PDB code: 2JYK) (Masliah et al. 2008) (d) the crystal structure of a nucleic acid four-way junction (PDB code: 1EGK) (Nowakowski et al. 2000) (e) high-resolution Dickerson-Drew Dodecamer B-DNA (PDB code: 4C64) (Lercher et al. 2014). The reduced representation was then subjected to the back-mapping algorithm. The obtained all-atom structure was refined as explained in the back-mapping procedure.

As the next target to test the algorithm a ‘real-case’ scenario was used. The DNA duplex structure (PDB code: 9BNA) (Westhof 1987). The DNA duplex chains were separated by the distance of 7 Å and each chain was extended. Multiplexed replica-exchange molecular dynamics (MREMD) (Czaplewski et al. 2009) was performed with a temperature range 200–350 K with 10 K intervals and two replicas in each temperature (32 trajectories per system in total). For each trajectory, 60 mln steps with time step 0.498fs were performed. The Berendsen thermostat was used with a coupling time of 49.8 fs. The exchange was attempted every 20000 steps. After simulation bin-less weighted histogram analysis (Kumar et al. 1992; Czaplewski et al. 2009) was performed to obtain different ensemble properties, and finally clustering at 220K was performed to obtain five dominant clusters. For the all-atom structure reconstruction, only the most probable cluster was used. The all-atom structure was reconstructed and refined as explained in the earlier section.

Data analysis

After recovering all-atom representation, the structure was subject to performance analysis. The total RMSD, by root-mean-square deviation per residue (root-mean-square fluctuations, RMSF), was computed, Euclidian RMSD (eRMSD)

(Bottaro et al. 2014), which indicates the similarity of contact maps between two structures, hydrogen bonds pattern, and the distribution of the torsional angles. The angles (α - χ) definition is shown in Fig. 6 (Frellsen et al. 2009). The eRMSD and the relevant dihedral angles were computed with Barnaba software (Bottaro et al. 2019). The angles' probability was computed with a bin size of 15°. The hydrogen bond pattern was computed with cpptraj with default options.

Results and discussion

In Table 1, the RMSD and the eRMSD of the back-mapped structures before and after minimization are shown. In all cases when the idealized coarse-grained structure is used the RMSD of the back-mapped structure is below 1.5 Å and the eRMSD is below 0.9. This indicates the high accuracy of the back-mapping algorithm. In the case of the simulated association of the DNA, the RMSD is less than 4 Å for the dominant cluster.

The RNA pseudoknot experimental structure (Fig. 7A) was converted into coarse-grain representation (Fig. S1B) and then back-mapped to all-atom representation (Fig. S1C). The structure has some clashes which arise from using the limited library of rotamers - only one rotamer available (Fig. S1C). A short minimization leads to the removal of all clashes and the final RMSD of the all-atom structure is 0.90 Å (Fig. 7B) which is a very good result as the back-mapping algorithm recovers all-atom structure from a low-resolution one.

The largest differences are observed in the loop-forming triplex structure (Fig. S2). The 14 native H-bonds and

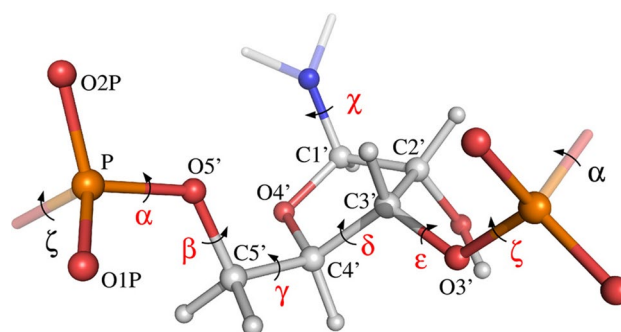
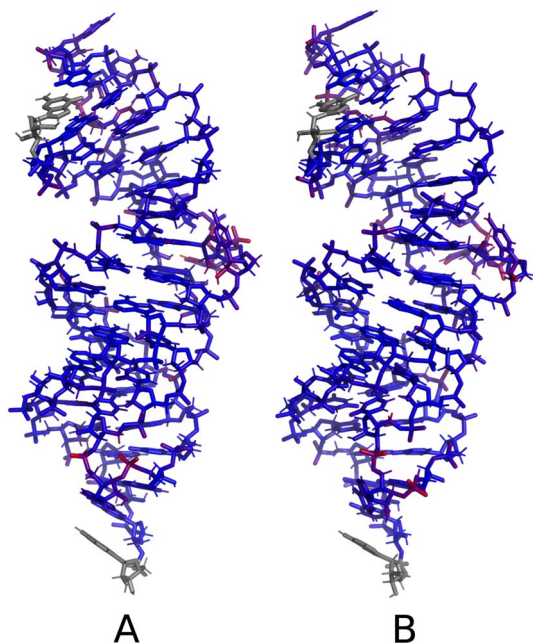


Fig. 6 The seven relevant dihedral angles in the central nucleotide (α to ζ) are indicated with red labels. Each label is placed on the central bond of the four consecutive atoms that define the dihedral angle. The χ angle describes the rotation of the base relative to the RNA backbone, while the six other angles define the course of the backbone. All atoms in the central nucleotide are labeled and colored according to atom type (oxygen: red, phosphor: yellow, nitrogen: blue, and carbon/hydrogen: grey). For clarity, the base is only partly shown. Reproduced from Frellsen et al. (2009)

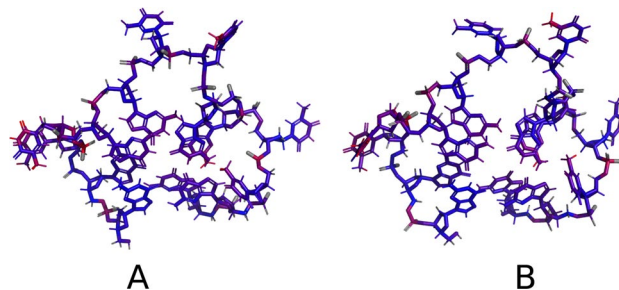
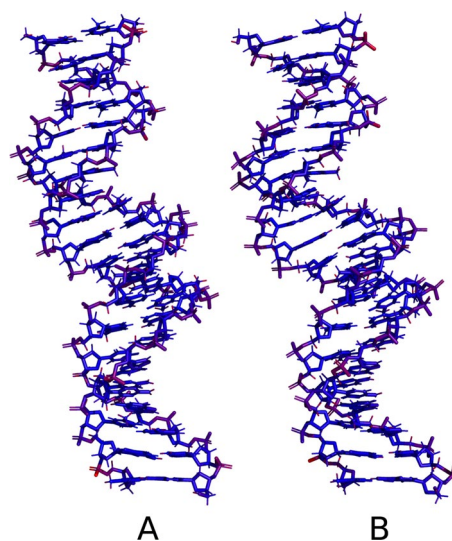
Table 1 RMSD(Å) and eRMSD for structures reconstructed before and after minimization

System	PDB code	RMSD before min	eRMSD before min	RMSD after min	eRMSD after min
RNA pseudoknot	2M8K	0.71	0.79	0.90	0.65
222 G-quadruplex structure	–	1.27	0.85	1.35	0.87
21 bp DNA duplex	2JYK	0.62	0.63	0.75	0.22
Nucleic acid four-way junction	1EGK	0.69	0.68	0.85	0.49
Dickerson-Drew Dodecamer B-DNA	4C64	0.51	0.38	0.61	0.18
B-DODECAMER D (MREMD)	9BNA	3.54	1.34	3.90	1.54

**Fig. 7** Comparison of the RNA pseudoknot experimental structure (A) with structure converted to coarse-grain representation and reconstructed by NARall software after short refinement (B). The coloring is blue to red based on the RMSD of each residue

19 non-native H-bonds were observed, with the eRMSD value of 0.65. This indicates that the contact pattern is preserved, though the H-bonds for triplex structures are not well preserved and the short structure optimization leads to the occurrence of more hydrogen bonds than in the native structure.

The 222-parallel G4 quadruplex (Fig. 8A) was converted into coarse-grained representation (Fig. S3B) and then back-mapped to all-atom representation (Fig. S3C) and refined with all-atom minimization (Fig. 8B). The final structure after short minimization revealed a low RMSD of 1.35Å revealing the largest deviation in the loop region (Fig. S4). The short refinement improved the quality of the backbone conformation (Fig. S3C and D). None of the hydrogen bonds were correctly reproduced, though the 5 out of 6 hydrogen bonds had either proper donor or acceptor

**Fig. 8** Comparison of the 222 G-quadruplex structure (A) with structure converted to coarse-grain representation and reconstructed by NARall software after short refinement (B). The coloring is blue to red based on the RMSD of each residue**Fig. 9** Comparison of the NMR Structure of a 21 bp DNA duplex preferentially cleaved by Human Topoisomerase II (PDB code: 2JYK) (A) with structure converted to coarse-grain representation and reconstructed by NARall software after short refinement (B). The coloring is blue to red based on the RMSD of each residue

atoms, with eRMSD of 0.87. This indicates that even though the h-bonds are reproduced correctly, the rotamers in the case of quadruplex require improvements (Fig. 8B).

The NMR Structure of a 21 bp DNA duplex preferentially cleaved by Human Topoisomerase II (Fig. 9A) was converted into coarse-grained representation (Fig. S5B) and then back-mapped to all-atom representation (Fig. S5C) and refined with all-atom minimization (Fig. 9B). The final structure after short minimization has an RMSD of 0.75 Å which is significantly lower than the RMSD of the experimental structure ensemble (4.03 Å). The largest deviations are observed at C3' and C5' ends (Fig. S6), especially in the backbone position of the phosphate groups (Fig. 9B). The eRMSD is 0.224 revealing that all contacts are preserved, moreover, 40 out of 40 h-bond obtained after structure recovery are present in the native structure (0 non-native h-bonds), however, 9 h-bonds were missing.

The crystal structure of a nucleic acid four-way junction, which is an RNA/DNA hybrid (Fig. 10A) was converted into a coarse-grained representation (Fig. S7B) and then back-mapped to all-atom representation (Fig. S7C) and refined with all-atom minimization (Fig. 10D). The structure after recovery has 0.85 Å RMSD to the native crystal structure, which is significantly below the crystal structure resolution (3.1 Å). It should be noted that there is no direct relation between the RMSD of the prediction or the back-mapped structure to the crystal structure resolution. However, both measurements indicate the quality of the prediction. The contacts between the bases are similar to the crystal structure with the eRMSD of 0.497. The rebuilt structure has 78 h-bonds (61 native and 17 non-natives) and 20 h-bonds that are present in the crystal structure are missing in the rebuilt structure. The RNA is rebuilt more accurately than the DNA (Fig. S8).

The ultra-high resolution Dickerson dodecamer (CGC GAATTCGCG; PDB code: 4C64) was converted to coarse-grained representation (Fig. S9B) and back-mapped with the NARall algorithm (Fig. S9C). After back-mapping

the RMSD is 0.51 Å, while the eRMSD is 0.38 (Fig. S9), whereas after short minimization the RMSD increases to 0.61 Å, while eRMSD drops to 0.18 (Fig. 11). The RMSF plots reveal the uniform distribution of quality of back-mapping along the chain without any significant raise at the ends (Fig. S10). Despite very low RMSD, out of 26 h-bonds present in the native structure: 15 h-bonds were predicted correctly with 11 h-bonds missing and 3 h-bonds predicted incorrectly.

We analyzed the angle distribution of the four rebuilt structures and compared them with experimentally determined (Fig. 12). The angles α and γ reveal a single maximum in the experimental structures while in our rebuilt structures those reveal two maxima with one overlapping with the experimental and one maximum shifted. The β angles are predicted correctly as the distributions overlap well. The δ angles have bimodal distribution both in experimental and in our rebuilt structures. One maximum (-75° angle) overlaps well while the second overlap only partially with angles shifted by approximately 30° . The ϵ angle distribution reveals a single maximum in the experimental structures while our rebuilt structures reveal two maxima, however, those two maxima overlap well with the experimental distribution. In the case of the ζ angles, the distribution reveals two well-separated maxima, while in our rebuilt structures those maxima are not well separated leading to an excessively narrow distribution of this angle. In the case of χ angle distribution, the two maxima revealed are similar in our rebuilt structures but are shifted by approximately 15° .

In the case of DNA duplex folding the dominant structure obtained after clustering is a properly paired DNA duplex (Fig. 13B). However, the NARES-2P force field has the tendency to produce excessively stretched structures (the 5' to 3' distance is too large) in comparison with the native duplex (Fig. 13A). This leads to multiple clashes after the recovery

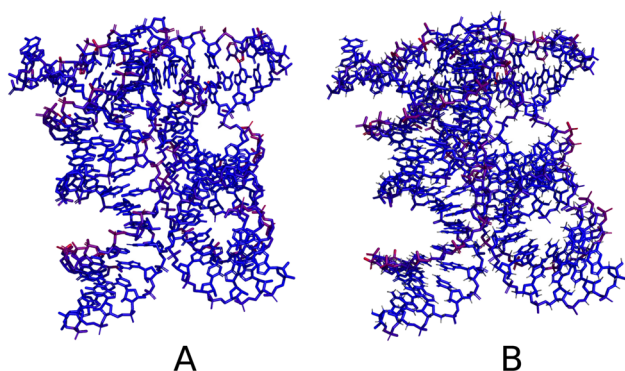


Fig. 10 Comparison of the crystal structure of a nucleic acid four-way junction (PDB code: 1EGK) (A) with structure converted to coarse-grain representation and reconstructed by NARall software after short refinement (B). The coloring is blue to red based on the RMSD of each residue

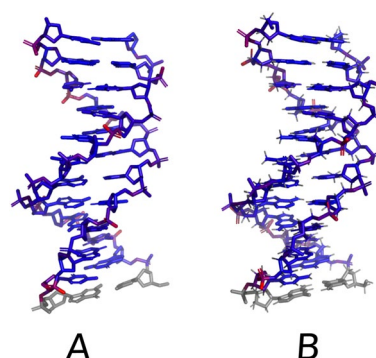


Fig. 11 Comparison of the crystal structure of a Dickerson dodecamer (PDB code: 4C64) (A) with structure converted to coarse-grained structure and then obtained from NARall after short refinement (B). The coloring is blue to red based on the RMSD of each residue

Fig. 12 The α , γ (A), β , δ (B) ϵ , ζ (C), and χ (D) angles distribution for the experimentally determined structures and the rebuilt structures with our algorithm for four chosen nucleic acids

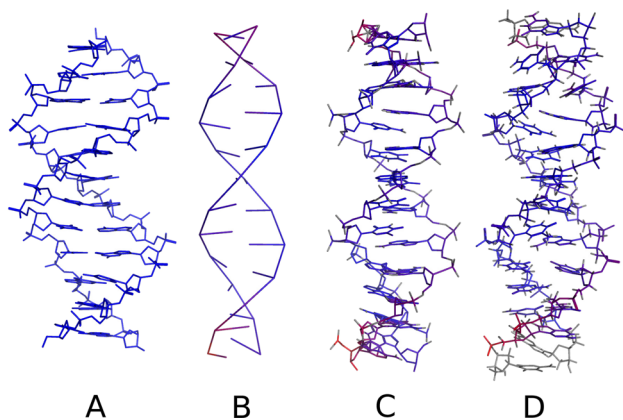
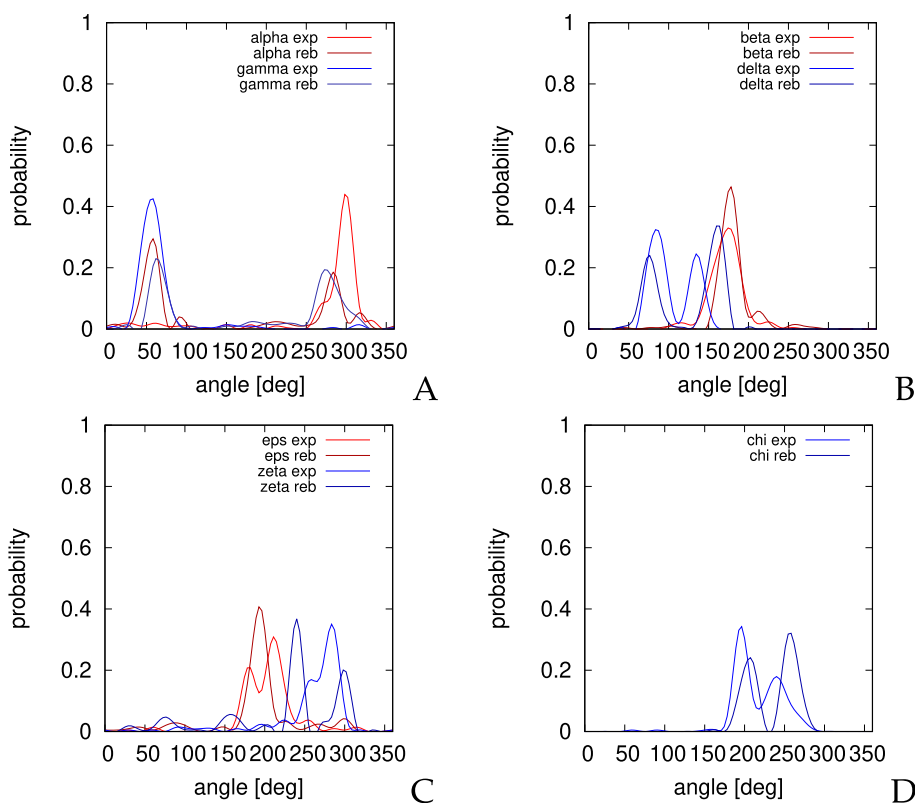


Fig. 13 Comparison of the DNA duplex experimental structure (A) with coarse-grain dominant structure obtained after MREMD simulation (B), reconstructed coarse-grained structure directly from NARall software (C) and after short refinement (D). The coloring is blue to red based on the RMSD of each residue

of the all-atom structure (Fig. 13C) as the base pairs are too close to each other. Those clashes are easily removed by the short refinement (Fig. 13D). Despite starting from separated chains and simulating folding, the final obtained structure after refinement has a low RMSD of 3.90Å. The largest deviations are observed in the terminal fragments of the chains (Fig. S11). The rebuilt structure has 17 native hydrogen bonds and 4 non-native hydrogen bonds, 11 hydrogen

bonds were missing, while the eRMSD value is 1.54. This indicates that the ab initio folded structure has most of the contacts and local structures predicted correctly.

Conclusions

In this article, we presented a new algorithm to back map NARES-2P coarse-grained structures to all-atom structures. Our back-mapping algorithm was applied to both idealized structures and real-case scenarios. This algorithm enables recovering all-atom details with good precision, however, the result depends on the quality of the NARES-2P model and the topology of the system. Our algorithm enables studying DNA/RNA dynamics in longer time ranges with only mild loss of resolution. Finally, as we now have a tool to recover the all-atom structure of the NARES-2P, we can take part in the RNA-puzzles experiment (Cruz et al. 2012) and we are currently taking part in the ongoing CAPRI experiment which has a protein-nucleic acids interface. The most of atomistic details are preserved as the RMSD and eRMSD to the reference structures are low, but the precision of h-bond recovery could be improved, we plan to improve this algorithm by introducing nucleobase-rotamers library dependence of the backbone conformation and improve the angle distribution mainly for α and γ angles. Our method should be handled with care when applied to study hydrogen

bonds pattern, and when the conformation of the backbone is essential for the non-helical structures.

The software is available from www.unres.pl webpage (NARall software).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11696-022-02634-x>.

Acknowledgements This work was supported by PRELUDIUM grant 2017/27/N/ST4/01907 from the National Science Center (Poland). Computational resources were also provided by (a) the supercomputer resources at the Informatics Center of the Metropolitan Computer Network (CI TASK) in Gdańsk at the Interdisciplinary Center of Mathematical and Computer Modeling, and (b) in-house 800-core Beowulf cluster at the Faculty of Chemistry, University of Gdańsk.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bottaro S, Bussi G, Pinamonti G, Reißer S, Boomsma W, Lindorff-Larsen K (2019) Barnaba: software for analysis of nucleic acid structures and trajectories. *RNA* 25(2):219–231. <https://doi.org/10.1261/rna.067678.118>
- Bottaro S, Di Palma F, Bussi G (2014) The role of nucleobase interactions in RNA structure and dynamics. *Nucl Acids Res* 42(21):13306–13314. <https://doi.org/10.1093/nar/gku972>
- Case DA, Betz R, Cerutti D, Cheatham Iii T, Darden T, Duke R, Giese T, Gohlke H, Goetz A, Homeyer N et al (2016) Amber 2016. University of California, San Francisco, p 810
- Cash DD, Cohen-Zontag O, Kim N-K, Shefer K, Brown Y, Ulyanov NB, Tzfati Y, Feigon J (2013) Pyrimidine motif triple helix in the *kluyveromyces lactis* telomerase rna pseudoknot is essential for function in vivo. *Proc Natl Acad Sci* 110(27):10970–10975. <https://doi.org/10.1073/pnas.1309590110>
- Cruz JA, Blanchet M-F, Boniecki M, Bujnicki JM, Chen S-J, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, Huang L, Lavender CA, Lisi V, Major F, Mikolajczak K, Patel DJ, Philips A, Puton T, Santalucia J, Sijenyi F, Hermann T, Rother K, Rother M, Serganov A, Skorupski M, Soltysinski T, Sripakdeevong P, Tuszyńska I, Weeks KM, Waldsich C, Wildauer M, Leontis NB, Westhof E (2012) Rna-puzzles: a casp-like evaluation of rna three-dimensional structure prediction. *RNA* 18(4):610–625. <https://doi.org/10.1261/rna.031054.111>
- Czaplewski C, Kalinowski S, Liwo A, Scheraga HA (2009) Application of multiplexing replica exchange molecular dynamics method to the unres force field: tests with α and $\alpha + \beta$ proteins. *J Chem Theory Comput* 5:627–640. <https://doi.org/10.1021/ct800397z>
- Dhamankar S, Webb MA (2021) Chemically specific coarse-graining of polymers: methods and prospects. *J Polym Sci* 59(22):2613–2643. <https://doi.org/10.1002/pol.20210555>
- Frellsen J, Moltke I, Thimm M, Mardia KV, Ferkinghoff-Borg J, Hamelryck T (2009) A probabilistic model of rna conformational space. *PLoS Comput Biol* 5(6):1–11. <https://doi.org/10.1371/journal.pcbi.1000406>
- Gopal SM, Mukherjee S, Cheng Y-M, Feig M (2010) Primo/primona: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins Struct Funct, Bioinf* 78(5):1266–1281. <https://doi.org/10.1002/prot.22645>
- Hang J, Wan R, Yan C, Shi Y (2015) Structural basis of pre-mrna splicing. *Science* 349(6253):1191–1198. <https://doi.org/10.1126/science.aac8159>
- He Y, Liwo A, Scheraga HA (2015) Optimization of a nucleic acids united-residue 2-point model (nares-2p) with a maximum-likelihood approach. *J Chem Phys* 143(24):243111. <https://doi.org/10.1063/1.4932082>
- He Y, Maciejczyk M, Ołdziej S, Scheraga HA, Liwo A (2013) Mean-field interactions between nucleic-acid-base dipoles can drive the formation of a double helix. *Phys Rev Lett* 110:098101. <https://doi.org/10.1103/PhysRevLett.110.098101>
- Hinckley DM, Freeman GS, Whitmer JK, de Pablo JJ (2013) An experimentally-informed coarse-grained 3-site-per-nucleotide model of dna: Structure, thermodynamics, and dynamics of hybridization. *J Chem Phys* 139(14):144903
- Khalili M, Liwo A, Jagielska A, Scheraga HA (2005) Molecular dynamics with the united-residue model of polypeptide chains. ii. langevin and berendsen-bath dynamics and tests on model α -helical systems. *J Phys Chem B* 109:13798–13810. <https://doi.org/10.1021/jp058007w>
- Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J Comput Chem* 13:1011–1021. <https://doi.org/10.1002/jcc.540130812>
- Lercher L, McDonough MA, El-Sagheer AH, Thalhammer A, Kriauconis S, Brown T, Schofield CJ (2014) Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem Commun* 50:1794–1796
- Li S, Olson WK, Lu X-J (2019) Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucl Acids Res* 47(W1):W26–W34
- Liwo A, Baranowski M, Czaplewski C, Gołaś E, He Y, Jagieła D, Krupa P, Maciejczyk M, Makowski M, Mozolewska MA, Niadzedvtski A, Ołdziej S, Scheraga HA, Sieradzan AK, Ślusarz R, Wirecki T, Yin Y, Zaborowski B (2014) A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J Mol Model* 20(8):2306. <https://doi.org/10.1007/s00894-014-2306-5>
- Liwo A, Czaplewski C, Sieradzan AK, Lipska AG, Samsonov SA, Murarka RK (2021) Theory and practice of coarse-grained molecular dynamics of biologically important systems. *Biomolecules* 11(9):1347. <https://doi.org/10.3390/biom11091347>
- Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *J Chem Theory Comput* 11(8):3696–3713 (PMID: 26574453)
- Masliah G, René B, Zargarian L, Fermandjian S, Mauffret O (2008) Identification of intrinsic dynamics in a DNA sequence preferentially cleaved by topoisomerase ii enzyme. *J Mol Biol* 381(3):692–706. <https://doi.org/10.1016/j.jmb.2008.06.024>
- Mazur AK (2011) Local elasticity of strained DNA studied by all-atom simulations. *Phys Rev E* 84:021903. <https://doi.org/10.1103/PhysRevE.84.021903>

- Mrázíková K, Mlýnský V, Kührová P, Pokorná P, Kruse H, Krepl M, Otyepka M, Banáš P, Šponer J (2020) UUCG RNA tetraloop as a formidable force-field challenge for md simulations. *J Chem Theory Comput* 16(12):7601–7617. <https://doi.org/10.1021/acs.jctc.0c00801>
- Nowakowski J, Shim PJ, Stout C, Joyce GF (2000) Alternative conformations of a nucleic acid four-way junction. *J Mol Biol* 300(1):93–102. <https://doi.org/10.1006/jmbi.2000.3826>
- Ouldrige TE, Louis AA, Doye JPK (2010) DNA nanotweezers studied with a coarse-grained model of DNA. *Phys Rev Lett* 104:178101. <https://doi.org/10.1103/PhysRevLett.104.178101>
- Palermo G, Casalino L, Magistrato A, Andrew McCammon J (2019) Understanding the mechanistic basis of non-coding rna through molecular dynamics simulations. *J Struct Biol* 206(3):267–279. <https://doi.org/10.1016/j.jsb.2019.03.004>
- Peter C, Kremer K (2009) Multiscale simulation of soft matter systems - from the atomistic to the coarse-grained level and back. *Soft Matter* 5:4357–4366. <https://doi.org/10.1039/B912027K>
- Poblete S, Bottaro S, Bussi G (2017) A nucleobase-centered coarse-grained representation for structure prediction of RNA motifs. *Nucl Acids Res* 46(4):1674–1683. <https://doi.org/10.1093/nar/gkx1269>
- Poblete S, Bottaro S, Bussi G (2018) Effects and limitations of a nucleobase-driven backmapping procedure for nucleic acids using steered molecular dynamics. *Biochem Biophys Res Commun* 498(2):352–358. <https://doi.org/10.1016/j.bbrc.2017.12.057>
- Sambriski E, Schwartz D, de Pablo J (2009) A mesoscale model of DNA and its renaturation. *Biophys J* 96(5):1675–1690. <https://doi.org/10.1016/j.bpj.2008.09.061>
- Shimizu M, Takada S (2018) Reconstruction of atomistic structures from coarse-grained models for protein-DNA complexes. *J Chem Theory Comput* 14(3):1682–1694 (PMID: 29397721)
- Sieradzan AK, Golon L, Liwo A (2018) Prediction of dna and rna structure with the nares-2p force field and conformational space annealing. *Phys Chem Chem Phys* 20:19656–19663. <https://doi.org/10.1039/C8CP03018A>
- Sieradzan AK, Krupa P, Wales DJ (2017) What makes telomeres unique? *J Phys Chem B* 121(10):2207–2219. <https://doi.org/10.1021/acs.jpcc.6b08780>
- Uusitalo JJ, Ingólfsson HI, Akhshi P, Tieleman DP, Marrink SJ (2015) Martini coarse-grained force field: extension to DNA. *J Chem Theory Comput* 11(8):3932–3945. <https://doi.org/10.1021/acs.jctc.5b00286>
- Wassenaar TA, Pluhackova K, Böckmann RA, Marrink SJ, Tieleman DP (2014) Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J Chem Theory Comput* 10(2):676–690. <https://doi.org/10.1021/ct400617g>
- Westhof E (1987) Re-refinement of the b-dodecamer d(cgcgaattcgcg) with a comparative analysis of the solvent in it and in the z-hexamer d(5brcg5brcg5brcg). *J Biomol Struct Dyn* 5(3):581–600. <https://doi.org/10.1080/07391102.1987.10506414>
- Wilson DM, Thompson LH (1997) Life without DNA repair. *Proc Natl Acad Sci* 94(24):12754–12757. <https://doi.org/10.1073/pnas.94.24.12754>
- Yan C, Hang J, Wan R, Huang M, Wong CCL, Shi Y (2015) Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* 349(6253):1182–1191
- Yoo J, Winogradoff D, Aksimentiev A (2020) Molecular dynamics simulations of DNA–DNA and DNA-protein interactions. *Curr Opin Struct Biol* 64:88–96. <https://doi.org/10.1016/j.sbi.2020.06.007>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.