Published online 13 March 2023

Nucleic Acids Research, 2023, Vol. 51, No. 6 **2633–2640**
https://doi.org/10.1093/nar/gkad136

Correlated motions in DNA: beyond base-pair step models of DNA flexibility

Kim López-Güell[1], Federica Battistini[1,2] and Modesto Orozco[1,2,*]

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri i Reixac 10-12, 08028 Barcelona, Spain and [2]Department of Biochemistry and Biomedicine, University of Barcelona, 08028 Barcelona, Spain

Received October 05, 2022; Revised February 08, 2023; Editorial Decision February 09, 2023; Accepted February 16, 2023

ABSTRACT

Traditional mesoscopic models of DNA flexibility use a reductionist-local approach, which assumes that the flexibility of DNA can be expressed as local harmonic movements (at the base-pair step level) in the helical space, ignoring multimodality and correlations in DNA movements, which have in reality a large impact in modulating DNA movements. We present a new multimodal-harmonic correlated model, which takes both contributions into account, providing, with a small computational cost, results of an unprecedented local and global quality. The accuracy of this method and its computational efficiency make it an alternative to explore the dynamics of long segments of DNA, approaching the chromatin range.

INTRODUCTION

DNA is a long and flexible polymer that has been typically represented by simplistic approaches such as the elastic rod or worm-chain-like models (1–3), which lack resolution and neglect sequence-dependent changes. The seminal work by Olson and Zhurkin (OZ; (4)) opened the possibility to describe DNA with sequence-dependence at base-pair step (bps) resolution. This model assumed that the deformation energy of DNA can be determined from that of individual bps following a simple harmonic model, where the energy of a full oligo is expressed in terms of a global stiffness matrix:

$$E\ (X) = \frac{1}{2}\ Y^T K Y\ = \frac{1}{2}\ K \Delta Y^2 \qquad (1)$$

where Y is a $6 \times N$ dimensional vector $Y = \{X_1, \ldots X_N\}$ with equilibrium values $Y = Y^0$, and the stiffness constant (K) is obtained by inverting a block covariance matrix (in the helical space), something that by construction implies the neglect of bps–bps correlations. That is global deformability is expressed exclusively in terms of local deformability by:

$$K = k_b T \mathbb{C}^{-1} = k_b T \begin{pmatrix} C_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_N^{-1} \end{pmatrix} \qquad (2)$$

where each bps (1 to N) defines a 6×6 (twist, roll, tilt, slide, rise and shift) stiffness matrix. In the original OZ work, the values were obtained from the analysis of the Protein Data Bank (PDB; (4)).

As described elsewhere (5), the practical use of the original OZ model was hampered by the low density of the bps ensembles of naked DNA available in the PDB. Fortunately, in another seminal work, Lankas and Langowski (LL; (6)) demonstrated that atomistic molecular dynamics (MD) simulations could be used to obtain dense ensembles, from which equilibrium values and stiffness matrices could be derived. Based on those ideas, the Ascona B-DNA consortium (ABC) performed a series of MD simulations to define deformability parameters for the ten unique bps of B-DNA (7,8). Even though the results were affected by the lack of accuracy of 2004 force-fields (9), they revealed the imprecision of the bps model and the need to consider the neighbouring steps, leading to a tetramer-based OZ model (7,8,10,11). This model implicitly introduces the 'static' correlation between bps, in which the equilibrium value of a given bps is dependent on the neighbouring ones, but not the dynamic correlation coupling the movements at two bps.

The analysis of the new generation of ABC simulations showed the existence of multimodal behaviour for some helical parameters, rising several warnings on the validity of the harmonic approximation implicit in equation 1 (10,11). This finding was received with some scepticism from the community due to the shortcomings of PARMBSC0 force-field (12–14), but were confirmed by simulations using a more accurate force filed PARMBSC1 (13,14), as well as by detailed analysis of experimental data (15–17). To account for bimodality, our group developed in 2020 a multimodal scheme (18,19), which uses the Gaussian finite mixture clustering (GFMC) (19,20) method to define the minimum number of unimodal distributions (typically between

*To whom correspondence should be addressed. Tel: +34 93 40 37156; Email: modesto.orozco@irbbarcelona.org

2 and 6 substates) that when combined reproduce the MD helical distribution for each tetramer. According to this approach the deformation energy estimate is computed by:

$$E\ (X) = \ -k_{\mathrm{b}}T\sum_{j\,=\,1}^{N}\ln\sum_{i\,=\,1}^{n}\mathrm{e}^{-\frac{1}{k_{\mathrm{B}}T}\left(\frac{1}{2}\Theta_{ij}\Delta\mathrm{X}_{ij}^{2}+E_{ij}\right)} \qquad (3)$$

where the first sum extends to the N bps, and the second from 1 to n number of substates in a given bps. E_{ij} is the relative energy of state i at bps j (shifting values between substates) and $\frac{1}{2}\Theta_{ij}\Delta X_{ij}^{2}$ is the harmonic term, which defines the deformation energy associated to substate i at bps j ($E_{\mathrm{harm}\ i,j}$). The equation above can be rewritten in a more compact manner as:

$$E\ (X) = \ -k_{\mathrm{b}}T\sum_{j\,=\,1}^{N}\ln\sum_{i\,=\,1}^{n}P_{ij}\mathrm{e}^{-\frac{1}{k_{\mathrm{B}}T}\left(E_{\mathrm{harm}\ i,j}\right)} \qquad (4)$$

where P_{ij} is the probability of substate i at bps j (obtained from the MD).

The multimodal formalism provides outstanding representations of the helical space at the bps level (18) but neglects the correlations between bps, i.e., the nearly instantaneous coupling between the movements at one bps and the neighbouring ones, which means that strong perturbations might be artifactually transferred along the duplex.

As described above, the lack of correlation was inherited from the original assumption of the block matrix nature of the global covariance model in OZ model. To our knowledge, Maddocks *et al.* (21) were the first to consider correlations by following the harmonic principle in equation 1, but considering that the covariance matrix is banded instead of a block matrix, which allowed them to derive an harmonic model that accounts for explicit correlations between neighbouring bps, but by construction neglects non-harmonic deformations. More recently, Zacharias's group (22) presented a method to solve a reduced version of the problem considered here, in particular, the coupling of the BI/BII bimodality and next neighbouring correlations, but the model is not extensible to all the other multimodalities.

We present here a new and general mesoscopic model that combines non-harmonic descriptions of DNA deformability with explicit consideration of correlations effects. This model, with a computational efficiency similar to that of previous mesoscopic models, provides local and global results of an unprecedented quality.

MATERIALS AND METHODS

The multimodal correlated model

Our previous local multimodal approach implies that each bps j samples a limited number of harmonic substates i, each of them with probability P_{ij}, and the total sampling at step j is just a weigthed sum of the harmonic samplings at each substate i (equation 4). In the limit of the uncorrelated multimodal approach (19,20), this concept can be extended by considering bp-bp steps defined by the join probability of sampling two substates (i and i') in two bps (j and j') and can be expressed simply as a product of probabilities

$$P_{ij}\cup\ P_{i'j'} = P_{ij}\ \cdot P_{i'j'} \qquad (5)$$

Unfortunately, as shown in this work, this equality is generally incorrect and the join probabilities $P_{ij}\cup P_{i'j'}$ needs to be determined. To this end we repeat the cluster annotation in (19), but obtaining clusters in overlapping 12×12 rather than 6×6 dimensional space, obtaining pentamer-dependent tetramer populations of substates $\{s_j^i\}$, each of them with an associated probability $\{p_j^i\}$ (see Supplementary Figure S1). Obtaining the combination of these probabilities is complex due to the $j-1\longleftrightarrow j\longleftrightarrow j+1$ dependence, which means that the probability of substate i at bps j is dependent simultaneously of the substate sampled at this specific conformation at $j-1$ (given by -1 pentamer) and that at $j+1$ (given by the + 1 pentamer). That is, the locality implicit to most of the mesoscopic models derived from OZ ideas is not maintained. We solved the problem by using a 1D Ising model (23), yielding to a set of states for the global duplex (I) defined by a set of bps substates for the N steps. Once selected the substates for each bps j, using Monte Carlo simulations, the effective Hamiltonian yielding the states as a combination of bps substates consistent with the expected population of correlated states can be obtained by:

$$E\ (I) = \ \gamma\sum_{j}\ln\left(\frac{P_j\cup P_{j+1}}{P_j\cdot P_{j+1}}\right)+\sum_{j}\ln(P_j) \qquad (6)$$

where $E(I)$ is the energy associated to duplex state I defined by $\{i^{(j\,=\,1)},\ldots,i^{(j\,=\,N)}\}$, where j stands for each of the N bps and substates $i\ \in I$ (at bps j), and γ is a real parameter, that can be used to modulate the direct and correlated effects (a value of 1.0 is used here; note that for $\gamma = 0$ the method converges to the uncorrelated multimodal scheme).

The Metropolis Monte Carlo procedure leads to a set of representative structures (in general, 10^3–10^4 structures are enough to provide good ensembles; (19)), each of such structures is defined by a given substate i at bps j obtained by explicitly considering correlation between substate populations at neighbouring steps. Note that the energy for the entire duplex can then be easily represented by:

$$E\,(X) = \sum_{I}\sum_{j=1}^{N}\frac{1}{2}\theta_j^{i,i'}\left(\Delta\,X_{\mathrm{j}}^{\mathrm{i,i'}}\right)^{2} \qquad (7)$$

where $\Theta_j^{i,i'}$ is a 12×12 banded stiffness matrix of the substate i at step j and substates $i'\ \in I$ (at bps j). In practice, the external elements out of the 6×6 block matrix are rather small for a given i, i' substates and a simple 6×6 stiffness matrix would be accurate enough, but for the sake of completeness 12×12 banded stiffness matrix (21) is considered in the model which allowed us to capture correlations inside a given state.

Mesoscopic samplings

Sampling of duplexes was obtained using Metropolis Monte Carlo algorithms (19,24) in the helical space. Approaches to transform these helical coordinates into Cartesian space have been previously developed (19) and recently improved to provide all Cartesian details from a helical coordinate ensemble (25).

Reference calculations

Atomistic MD calculations were used as a source of parameters as well as for validation purposes. Parametrization of the model was done from a variety of PARMBSC1 (13) trajectories covering all unique pentamers (14–18,26). Simulations 0.5–10 μs long were obtained using *state-of-the-art* protocols at constant pressure and temperature ($P = 1$ atm, $T = 300$ K). A series of additional trajectories (obtained using equivalent simulation conditions) were used for benchmarking and validation of the method, as well as for describing the nature and magnitude of correlation effects. Analysis were carried out using NAFlex, BIGNASim and Curves+ (26–28). Similarity analysis between trajectories were performed using Hess metrics as described elsewhere (1,2,29).

Persistence lengths were calculated using the SerraNA software (30). All trajectories are available at our BigNAsim database (14–18,26).

RESULTS

Local correlations in the helical space

We explored first the magnitude of the instantaneous correlations along the DNA using a 10 μs molecular dynamics trajectory of the Drew–Dickerson dodecamer (DDD; (14,15)), which we know reproduces well experimental duplex properties. Results in Figure 1 demonstrate that the assumption of no correlation between bps, when adapted to the tetramer description, is incorrect. Generally, the strongest couplings between neighbouring bps are homo-cross correlations (i.e. correlation between the same helical parameters at base step j and its neighbours) mainly in shift, tilt and twist movements (all of them negative), but some hetero-cross correlations (typically positive) are not negligible. Fortunately, correlations decay quickly with sequence, and they are small except for the nearest neighbour bps ($j \rightarrow j + 1$; see Figure 1A). Very interestingly, couplings are not uniform along the duplex (Figure 1B), suggesting a sequence-dependent correlation pattern.

Analysis of time-convergence in the cross-correlation indexes shows that even in those cases with strong correlations (see selected cases in Supplementary Figure S2), results converge in approximately 10 ns. Comparison of 1 and 10 μs simulations failed to detect any significant difference in correlation pattern (data not shown), confirming that cross-correlations are not artifacts of limited sampling. Furthermore, by using sliding time windows, we show that cross-correlation disappear after ∼100 ps, indicating that they are coupled to fast relaxation movements of the fiber (see Supplementary Figure S3).

To check for the universality of our findings we extended the analysis to the mini-ABC dataset (18), a series of 13 MD simulations of 18-mer duplexes containing all the 136 unique tetramers. As suggested by DDD results above there is a fast decay of correlations along sequence distance (Supplementary Figure S4), and couplings are mainly linked to shift, tilt and twist movements. Sparse cross-correlations matrices appear (Supplementary Figure S5), with the predominance (as DDD simulations suggested) of shift, tilt and twist couplings. Strong sequence dependence shows up, with higher correlations, when involving bps showing multimodality (Supplementary Figure S5). For example, tetramers with CpG central steps show large complexity in the correlation maps, while movements in ApT bps are quite uncorrelated from their neighbours. Finally, a significant directionality is evident in the couplings, reflecting the pentamer dependence, and compensatory effects between homo (typically negative) and hetero (typically positive) cross-correlations are evident, as already suggested by results obtained for DDD (see Figure 1 and Supplementary Figure S4).

The impact of multimodality and correlations in the global structure and flexibility of DNA duplex: the correlated multimodal model

As described above, multimodal methods provide accurate representations of the bps conformational space (see examples in columns UM and CM in Figure 2), something that by construction, cannot be achieved by harmonic models (columns OZ and CH in Figure 2), which reduce the sampled space and might suggest structures that are rarely sampled in the atomistic simulations (column MD in Figure 2) as the most stable geometries for the bps. Note that, not major differences are found at the intra-bps helical space between correlated and non-correlated multimodal schemes (columns UM and CM in Figure 2). The neglect of correlation effects introduces significant errors in the helical distributions beyond the tetramer, which are very evident looking at the homo-helical distribution maps between steps j and $j + 1$ (Figure 3 compare columns MD with OZ, UM). Clearly ignoring the coupling between the movements of neighbouring steps can lead to combinations of substates that are forbidden by the physics of DNA.

A simple harmonic correlated model (column CH in Figures 2 and 3) captures well the general shape of the inter-bps distributions but fails to capture the details of the multimodal distribution at both the inter-bps and intra-bps level. The multimodal correlated model presented here is not only able to reproduce well the MD distribution at the intra-bps level (Figure 2 column CM) but also matches the MD inter-bps distributions (Figure 3, column CM).

There is not enough experimental data to corroborate the goodness of the distributions shown in Figures 2 and 3, as it would require a dense population of unperturbed tetramers and pentamers in databases. However, we were able to extract a few cases from the PDB with enough data to show that the distributions predicted by our correlated multimodal model are correct (see Supplementary Figure S6), providing extra support to our new theoretical approach.

The lack of correlation in traditional models leads to the accumulation of large helical changes in neighbouring steps, leading to an artefactual increase in pentamer flexibility, which is propagated along the entire duplex. This overestimation is visible in the distribution of the end-to-end distance, the magnitude of the first eigenvalues associated to the most important deformation of DNA and the global bending (see Figure 4A–D), as well as in different non-local helical descriptors (see Supplementary Figure S7). This problem is largely corrected by the multi-

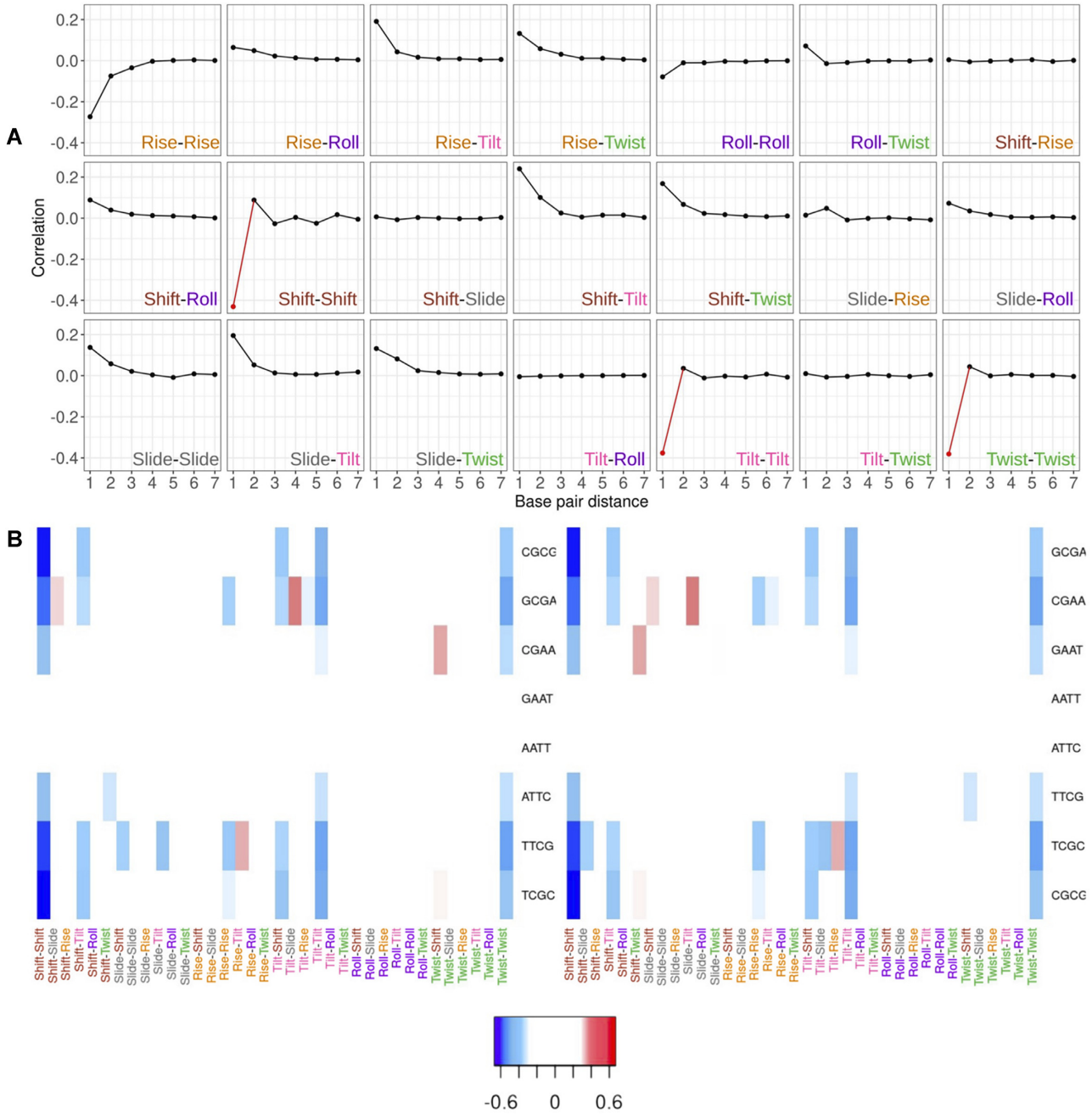

Figure 1. Correlation effects in DDD simulation. (**A**) Correlation between helical coordinates with the distance (point with $|r| > 0.3$ are shown in red). (**B**) Heat maps showing $j \leftrightarrow j+1$ (left panel) and $j \leftrightarrow j-1$ (right panel) correlations for each possible bps and parameter combination along the DDD.

modal correlated functional, which not only reproduces the magnitude (Figure 4A) but also the nature of the essential deformation modes (Figure 4D). Better approximations to the MD values would require the introduction of correlation effects beyond the pentamer level (something which might be explored in the future), but the current slight deviation from MD is in fact desirable to correct a certain overestimation (approximately 20%) of the duplex stiffness in PARMBSC1 atomistic MD simulations. This is visible in the persistence length (PL) of the central 36mer derived from MD ensembles: 62.5 ± 1.2 nm (as computed in (31)), which compares with a PL equal to 49.5 ± 1.1 nm, determined using the same protocol, from the Monte Carlo ensembles using the multimodal correlated model. Note that the accepted experimental value for the persistence length of a DNA with a random sequence is approximately 50 nm, and the value obtained using the uncorrelated multimodal method is approximately 38 nm (see Supplementary Table S1).

To further validate the accuracy of this method, we evaluated the sequence-dependent recognition by papillomavirus E2 protein (22). Following Zacharias' protocol, we calcu-

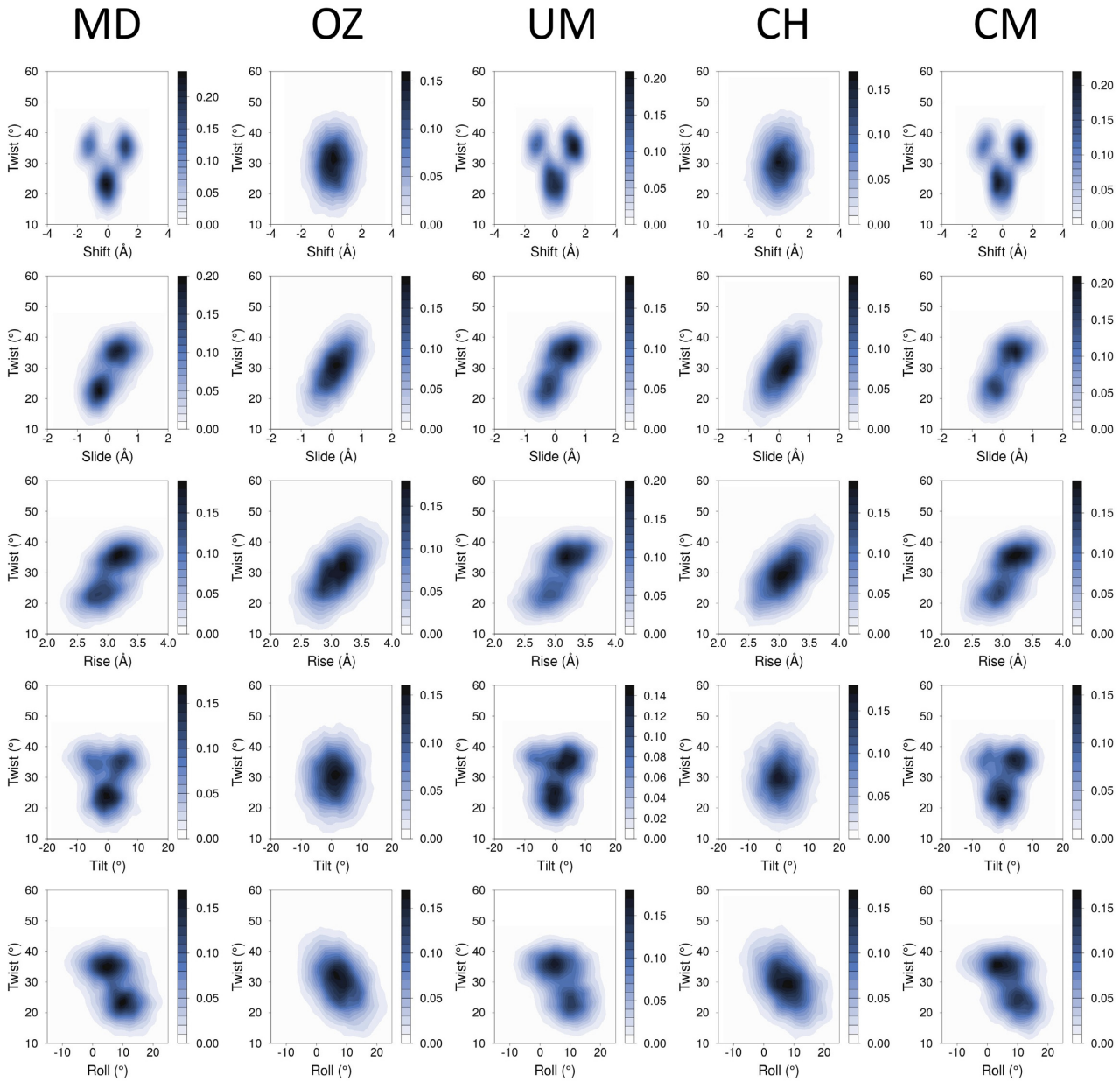

Figure 2. Intra-bps ensembles using different methods. Selected examples of the intra-bps ensembles (in helical space) of the tetramer d(TCGA) sampled using different methods: atomistic MD simulation (the reference) (MD), uncorrelated harmonic model (OZ), uncorrelated multimodal model (UM), correlated harmonic model (CH) and correlated multimodal model (CM) presented here.

lated the deformation energy between the structure predicted by our method and the DNA sequence bound to the protein (conformation achieved after a short MD simulation). We calculated the deformation energy for each 12-mer changing the central tetramer and comparing the deformation energy with the experimental one (32). Our model could predict the worst and the best sequence (see Supplementary Table S2), while the central ones have values very similar as in the Zacharias model. Furthermore, we compared the persistence length of 200 bs-long using structured calculated by our model and the experimental data (33). In detail we used poly(AA), poly(TA) and poly(CC) and we found a good correlation with experimental values (see Supplementary Table S3), being the poly(TA) the most flexible sequence and poly(AA) the stiffest.

DISCUSSION

We present here for the first time a comprehensive analysis of anharmonicity and correlated movements in B-DNA and the ability of our mesoscopic model to capture them. Available local models derived from the OZ (4) approach, despite their power, fail to reproduce the bps helical space at the tetramer level and are unable to capture the coupled movements at neighbouring bps. The local multimodal approach, previously developed in our group (19), improves dramat-

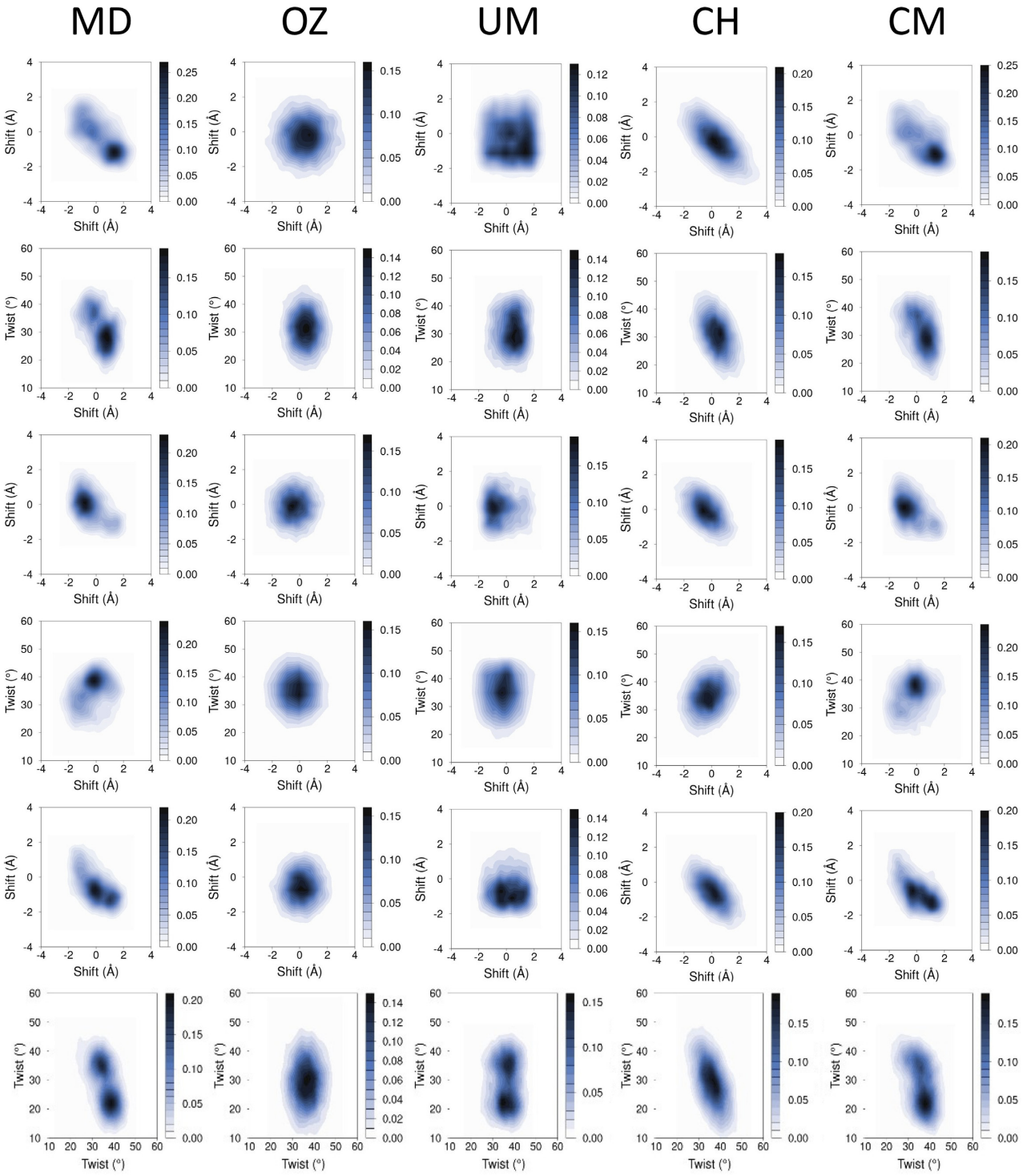

Figure 3. Inter-bps ensembles using different methods. Selected examples of the inter-bps parameters for the step *j* against the neighbouring step *j* + 1. Atomistic MD simulation (the reference) (MD), uncorrelated harmonic model (OZ), uncorrelated multimodal model (UM), correlated harmonic model (CH) and correlated multimodal model (CM) presented here. Pentamers represented correspond (from top to bottom and showing only the Watson strand): d(ATGGA), d(CTCTG), d(TTGGG), d(TCAAA) and d(TCGAG) and d(GCCGG).

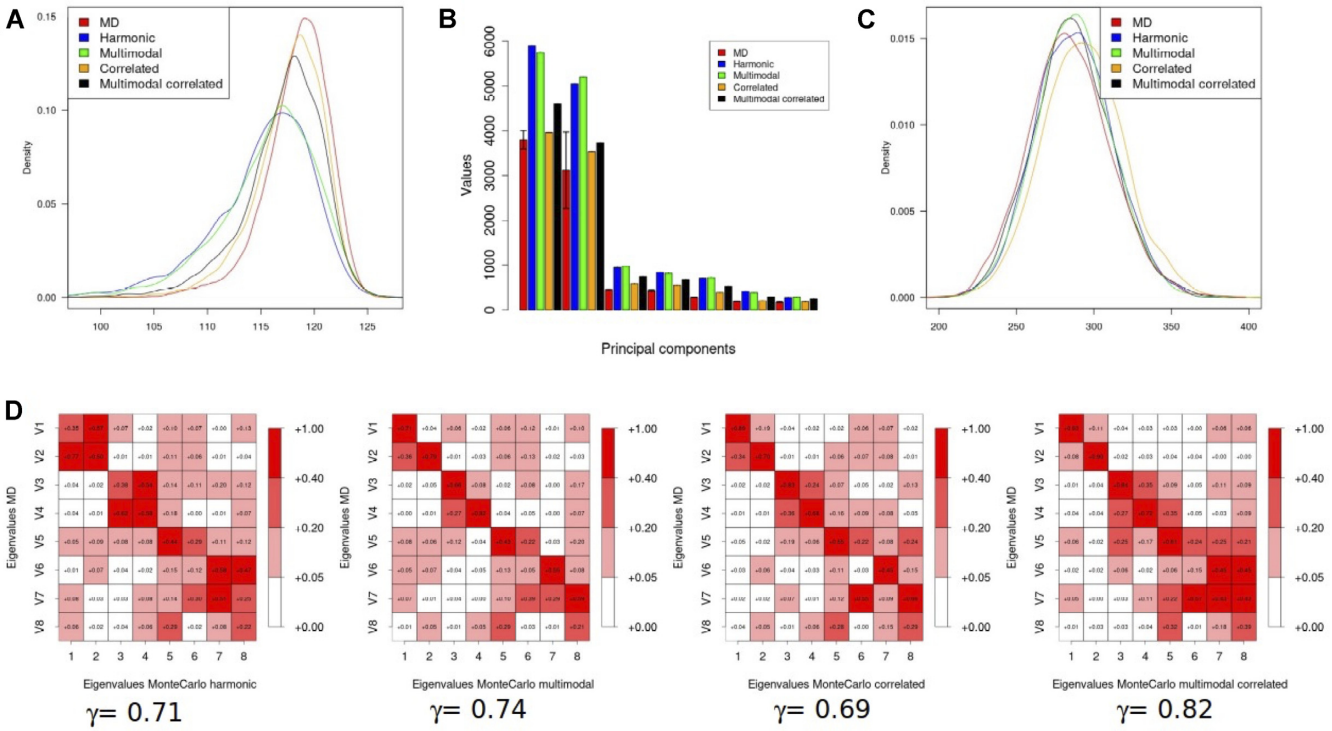

Figure 4. DNA properties and deformation comparisons. (**A**) Normalized distribution of end-to-end distance (in Å). (**B**) Eigenvalues associated with the first seven essential movements of the duplex (in $Å^2$; error bar in MD simulations obtained by comparing the eiganvalues in first and second halves of the trajectory). (**C**) Normalized distribution of global bending (computed using the roll and tilt contributions at each bps). (**D**) Eigenvector to eigenvector comparison of the essential deformation modes sampled by Monte Carlo with different Hamiltonian definition (from left to right: harmonic-uncorrelated, multimodal uncorrelated, harmonic correlated and multimodal correlated). The similarity Hess index (γ) is shown to summarize the similarity matrix (self-similarity obtained by comparing first and second halves of the MD trajectory is 0.95). All values correspond to the central 36- of a 40-mer duplex.

ically the ability to reproduce individual bps ensembles, but fails to reproduce dynamic correlations between neighbouring bps, leading to incorrect helical distributions at the pentamer level and to an overestimation of predicted duplex flexibility. The correlated harmonic model captures well the coupling between the movement of the neighbouring bps, but neither tetramer nor pentamer distributions could be reproduced correctly. Our new correlated multimodal approach can reproduce both local (at tetramer and pentamer level) and global flexibility. It can correct the errors at the base-pair step level, which appear in harmonic-based calculations, and, at the same time, is able to correct the overestimation of global DNA flexibility arising from calculations that neglect the correlation in the movement of neighbouring steps. The method has a reduced computational cost, which allows to push the limits of mesoscopic modelling of DNA structure and flexibility.

DATA AVAILABILITY

The program is available in the Zenodo repository DOI: 10.5281/zenodo.7628703.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are indebted to Prof. A. Noy for help with analysis and M. Orozco-Ruiz for help in setting the Ising model.

FUNDING

Center of Excellence for HPC H2020 European Commission; 'BioExcel – Centre of Excellence for Computational Biomolecular Research' [823830]; Spanish Ministry of Science [RTI2018-096704-B-100, PID2021-122478NB-I00]; Instituto de Salud Carlos III–Instituto Nacional de Bioinformatica, Fondo Europeo de DesarrolloRegional [ISCIII PT 17/0009/0007]; European Regional Development Fund, ERFD Operative Programme for Catalunya, the Catalan Government AGAUR [SGR2017-134]. The IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from the MINECO. Modesto Orozco is an ICREA Academy scholar. K.L.G. is a M4L student. The open access publication charge for this paper has been waived by Oxford University Press – *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.
Conflict of interest statement. None declared.

REFERENCES

1. Orozco,M., Pérez,A., Noy,A. and Luque,F.J. (2003) Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.*, **32**, 350–364.

2. Orozco,M., Noy,A. and Pérez,A. (2008) Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.*, **18**, 185–193.
3. Dans,P.D., Walther,J. and Gómez,H. (2016) Multiscale simulation of DNA. *Curr. Opin. Struct. Biol.*, **37**, 29–45.
4. Olson,W.K., Gorin,A.A., Lu,X.-J., Hock,L.M. and Zhurkin,V.B. (1998) DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proc. Natl. Acad. Sci.*, **95**, 11163–11168.
5. Pérez,A., Lankas,F., Luque,F.J. and Orozco,M. (2008) Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res.*, **36**, 2379–2394.
6. Lankas,F., Sponer,J., Langowski,J. and Cheatham,T.E. (2003) DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.*, **85**, 2872–2883.
7. Beveridge,D.L., Barreiro,G., Suzie Byun,K., Case,D.A., Cheatham,T.E., Dixit,S.B., Giudice,E., Lankas,F., Lavery,R., Maddocks,J.H. *et al.* (2004) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys. J.*, **87**, 3799–3813.
8. Dixit,S.B., Beveridge,D.L., Case,D.A., Cheatham,T.E. 3rd, Giudice,E., Lankas,F., Lavery,R., Maddocks,J.H., Osman,R. *et al.* (2005) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J.*, **89**, 3721.
9. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E., Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.
10. Lavery,R., Zakrzewska,K., Beveridge,D.L., Bishop,T.C., Case,D.A., Cheatham,T., Dixit,S., Jayaram,B., Lankas,F., Laughton,C. *et al.* (2010) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.*, **38**, 299–313.
11. Pasi,M., Maddocks,J.H., Beveridge,D., Bishop,T.C., Case,D.A., Cheatham,T., Dans,P.D., Jayaram,B., Lankas,F., Laughton,C. *et al.* (2014) ABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.*, **42**, 12272–12283.
12. Pérez,A., Luque,F.J. and Orozco,M. (2007) Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc.*, **129**, 14739–14745.
13. Ivani,I., Dans,P.D., Noy,A., Pérez,A., Faustino,I., Hospital,A., Walther,J., Andrio,P., Goñi,R., Balaceanu,A. *et al.* (2015) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55–58.
14. Dans,P.D., Ivani,I., Hospital,A., Portella,G., González,C. and Orozco,M. (2017) How accurate are accurate force-fields for B-DNA?*Nucleic Acids Res.*, **44**, gkw1355.
15. Dans,P.D., Pérez,A., Faustino,I., Lavery,R. and Orozco,M. (2012) Exploring polymorphisms in B-DNA helical conformations. *Nucleic Acids Res.*, **40**, 10668–10678.
16. Dans,P.D., Faustino,I., Battistini,F., Zakrzewska,K., Lavery,R. and Orozco,M. (2015) Unraveling the sequence-dependent polymorphic behavior of d (CpG) steps in B-DNA. *Nucleic Acids Res.*, **42**, 11304–11320.
17. Balaceanu,A., Pasi,M., Dans,P.D., Hospital,A., Lavery,R. and Orozco,M. (2017) The role of unconventional hydrogen bonds in determining BII propensities in B-DNA. *J. Phys. Chem. Lett.*, **8**, 21–28.
18. Dans,P.D., Balaceanu,A., Pasi,M., Patelli,A.S., Petkevičiūtė,D., Walther,J., Hospital,A., Bayarri,G., Lavery,R., Maddocks,J.H. *et al.* (2019) The static and dynamic structural heterogeneities of B-DNA: extending Calladine-Dickerson rules. *Nucleic Acids Res.*, **47**, 11090–11102.
19. Walther,J., Dans,P.D., Balaceanu,A., Hospital,A., Bayarri,G. and Orozco,M. (2020) A multi-modal coarse grained model of DNA flexibility mappable to the atomistic level. *Nucleic Acids Res.*, **48**, e29.
20. Day,N.E. (1969) Estimating the components of a mixture of normal distributions. *Biometrika*, **56**, 463.
21. Gonzalez,O., Petkevičiūtė,D. and Maddocks,J.H. (2013) A sequence-dependent rigid-base model of DNA. *J. Chem. Phys.*, **138**, 055102.
22. Liebl,K. and Zacharias,M. (2021) Accurate modeling of DNA conformational flexibility by a multivariate ising model. *Proc. Natl. Acad. Sci.*, **118**, e2021263118.
23. Ising,E. (1925) Beitrag zur theorie des Ferromagnetismus. *Zeitschrift Für Phys. 1925 311*, **31**, 253–258.
24. De Bruin,L. and Maddocks,J.H. (2018) cgDNAweb: a web interface to the cgDNA sequence-dependent coarse-grain model of double-stranded DNA. *Nucleic Acids Res.*, **46**, W5–W10.
25. Louison,K.A., Dryden,I.L. and Laughton,C.A. (2021) GLIMPS: a machine learning approach to resolution transformation for multiscale modeling. *J. Chem. Theory Comput.*, **17**, 7930–7937.
26. Hospital,A., Andrio,P., Cugnasco,C., Codo,L., Becerra,Y., Dans,P.D., Battistini,F., Torres,J., Goñi,R., Orozco,M. *et al.* (2016) BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res.*, **44**, D272–D278.
27. Hospital,A., Faustino,I., Collepardo-Guevara,R., Gonzalez,C., Gelpi,J.L. and Orozco,M. (2013) NAFlex: a web server for the study of nucleic acid flexibility. *Nucleic Acids Res.*, **41**, W47–W55.
28. Blanchet,C., Pasi,M., Zakrzewska,K. and Lavery,R. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.*, **39**, W68–W73.
29. Pérez,A., Blas,J.R., Rueda,M., López-Bes,J.M., de la Cruz,X. and Orozco,M. (2005) Exploring the essential dynamics of B-DNA. *J. Chem. Theory Comput.*, **1**, 790–800.
30. Velasco-Berrelleza,V., Burman,M., Shepherd,J.W., Leake,M.C., Golestanian,R. and Noy,A. (2020) SerraNA: a program to determine nucleic acids elasticity from simulation data. *Phys. Chem. Chem. Phys.*, **22**, 19254–19266.
31. Noy,A. and Golestanian,R. (2012) Length scale dependence of DNA mechanical properties. *Phys. Rev. Lett.*, **109**, 228101.
32. Kim,S.S., Tam,J.K., Wang,A.F. and Hegde,R.S. (2000) The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem.*, **275**, 31245–31254.
33. Geggier,S. and Vologodskii,A. (2010) Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 15421–15426.