Published online 17 May 2013 Nucleic Acids Research, 2013, Vol. 41, Web Server issue W47–W55
doi:10.1093/nar/gkt378

NAFlex: a web server for the study of nucleic acid flexibility

Adam Hospital[1,2,3,4], Ignacio Faustino[1,2,3], Rosana Collepardo-Guevara[1,2,3], Carlos González[5], Josep Lluis Gelpí[1,2,3,6,7,*] and Modesto Orozco[1,2,3,4,7,*]

[1]Institute for Research in Biomedicine (IRB Barcelona), Molecular Modelling and Bioinformatics (MMB) Department, 08028, Barcelona, Spain, [2]Joint IRB-BSC Programme on Computational Biology, Molecular Modelling and Bioinformatics (MMB) Department, 08028, Barcelona, Spain, [3]Barcelona Supercomputing Center, Life-Science Department, 08034, Barcelona, Spain, [4]National Institute of Bioinformatics, Structural Bioinformatics Node Division, 08028, Barcelona, Spain, [5]Instituto de Química Física 'Rocasolano' C.S.I.C., Biological Physical Chemistry Department, 28006, Madrid, Spain, [6]National Institute of Bioinformatics, Computational Bioinformatics Node Division, 08034, Barcelona, Spain and [7]Department of Biochemistry and Molecular Biology, University of Barcelona, Barcelona, Spain

Received January 30, 2013; Revised April 16, 2013; Accepted April 18, 2013

ABSTRACT

We present NAFlex, a new web tool to study the flexibility of nucleic acids, either isolated or bound to other molecules. The server allows the user to incorporate structures from protein data banks, completing gaps and removing structural inconsistencies. It is also possible to define canonical (average or sequence-adapted) nucleic acid structures using a variety of predefined internal libraries, as well to create specific nucleic acid conformations from the sequence. The server offers a variety of methods to explore nucleic acid flexibility, such as a colorless wormlike-chain model, a base-pair resolution mesoscopic model and atomistic molecular dynamics simulations with a wide variety of protocols and force fields. The trajectories obtained by simulations, or imported externally, can be visualized and analyzed using a large number of tools, including standard Cartesian analysis, essential dynamics, helical analysis, local and global stiffness, energy decomposition, principal components and *in silico* NMR spectra. The server is accessible free of charge from the mmb.irbbarcelona.org/NAFlex webpage.

INTRODUCTION

Nucleic acids are polymorphic molecules whose conformations depend not only on the sequence but also on external factors, such as temperature, solvent, ionic environment and presence of ligands. Sequence and environment-dependent polymorphisms make the experimental description of nucleic acids extremely difficult. Our coverage of the nucleic acid structural space in the near future is not likely to approach that already available for proteins (1–3). The experimental information regarding flexibility is even more scarce, being partial and very much limited to the B-type DNA duplex (1–3). Only recently, NMR spectroscopy has started to provide dynamical information at the atomistic level for some model nucleic acid systems (4–6). Although the results are of impressive quality and impact, the technique is clearly still unable to provide a complete description of the general flexibility of nucleic acids.

In the absence of an experimental approach, simulation techniques are now widely accepted tools with which to describe nucleic acid structure and flexibility (7–11). According to PUBMED, in 2012, nearly 900 articles were published quoting the words 'DNA', 'RNA' or 'nucleic acids' in combination with 'simulation' or 'molecular dynamics'. Given the improvements in the simulation methods and the accessibility to increasingly faster computers, greater popularization of the field is expected in the near future (9).

Theoretical approaches for the study of nucleic acids are diverse, but can be classified on the basis of two basic parameters: (i) the nature of the Hamiltonian used; and (ii) the level of resolution (7,8,11,12). The simplest simulation approaches, such as the wormlike-chain (WLC, 13) model, assume simple Hamiltonians that describe global helical properties of the DNA fiber. Such methods were developed to examine general properties of very long fibers of canonical B-DNA with great computational efficiency. Atomistic molecular dynamics (MD) takes a

*To whom correspondence should be addressed. Tel: +34 934037155; Fax: +34 934037157; Email: modesto.orozco@irbbarcelona.org. Correspondence may also be addressed to Josep Lluis Gelpí. Tel: +34 934037155; Fax: +34 934037157; Email: gelpi@ub.edu

© The Author(s) 2013. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

completely different approach. It combines a complex physical Hamiltonian with a rigorous method based on the resolution of Newton's equations of motion to obtain trajectories. MD allows an accurate atomistic description of any nucleic acid (8,9), but at the expense of a very large computational cost. The applicability of MD for very large systems or for the analysis of processes happening over large time scales is very limited. Between WLC and MD lie a wide variety of mesoscopic simulation methods. These were designed for the study of medium-to-large DNA fibers at the base-pair or even pseudo-base level of resolution using Hamiltonians of intermediate complexity (7,8,11–16).

For many years, the simulation of nucleic acids was performed by a small number of expert groups, often the developers of the software, force fields or algorithms. As the popularity of the simulation techniques has increased, groups with limited knowledge of the theoretical tools and often of the physics behind them have become users of these approaches. This development has generated significant confusion, as the potential user now faces a plethora of simulation packages based on a wide range of physical models and dealing with different levels of resolution. Furthermore, in general, a nonexpert user can find little information on how to setup the systems and encounters significant problems when attempting to perform meaningful analysis of the trajectories collected.

Here, we present NAFlex, a web tool designed to facilitate the use of nucleic acid simulation tools for newcomers to the field. The server allows the introduction of nucleic acid structures from diverse sources, as well as automatic structure generation from the sequences. The structural models can be subjected to several simulations, these are based on: (i) a colorless WLC model (13); (ii) a base-pair resolution mesoscopic model (14,15); or (iii) atomistic MD simulations. In the latter case, MDWeb technology (17) is used to help the user during all the setup and equilibration steps, providing all the input files required to launch the simulations. Finally, the trajectory obtained (or uploaded) can be visualized and analyzed using a large set of nucleic acid-specific tools. The server is freely accessible from the mmb.irbbarcelona.org/NAFlex webpage.

MATERIALS AND METHODS

NAFlex is divided (see Figure 1) into three main blocks: (i) *Input*; (ii) *Simulation engines*; and (iii) *Analysis*. The webserver has been designed to obtain a maximum coverage of potential user needs and as such is extremely flexible at the three levels (*input*, *simulation and analysis*).

Input

The user can introduce nucleic acid information in many ways depending on the nature of the problem and the calculation planned (see Figure 1). Thus, it is possible to upload a Protein Data Bank (PDB) structure (18) (of either isolated or complexed nucleic acids), or any user-derived structural model. The user can also upload a trajectory (all usual formats are accepted), which is then sent directly to the analysis modules of the server. When the user plans to work with canonical DNA and/or RNA duplexes, NAFlex allows the automatic generation of double helices using the Nucleic Acid Builder (nab) program from the Ambertools package (19). Structure generation can use parameters from various sources: (i) fiber diffraction data (20,21); (ii) sequence-dependent average X-ray structural information (1); (iii) tetramer-dependent average MD-simulation results (22–24); and (iv) user-defined helical values. The structures generated can be relaxed later to remove any distortion that may occur during model generation before being used in mesoscopic or atomistic simulations. In the case of WLC calculations, B-DNA geometry is assumed, and only sequence information is required as input.

After input, the server checks the structure to correct for potential gaps or missing atoms. Before launching the simulation, the setup procedures and a series of quality controls (see server help) are applied to the model. The server's check list includes, among others, alternate atom/residue locations, unusual distances between consecutive bases, steric clashes and the presence of metal ions or modified nucleotides/ligands. The server warns the user about potential structural errors (not trivial to correct).

Simulation engines

NAFlex offers the nonexpert user a selected variety of simulation tools that differ in complexity and resolution and that are designed to make their use as simple as possible.

WLC model (13) was introduced using the formalism of Jian *et al.* (25,26), implemented into a Monte Carlo sampling procedure. Accordingly, the DNA is represented as a set of N beads, each comprising M base-pairs (typically M = 10 base-pairs, but the exact bead resolution is selected by the user), and the potential energy of the DNA is defined by a simple Hamiltonian (see the help section for additional details):

$$E = E_S + E_B + E_T + E_{ele} \quad (1)$$

where the stretching energy (E_S) is computed as:

$$E_S = 0.5 K_S \sum_{i=1}^{N-1} (l_i - l_0)^2 \quad (2)$$

where K_s is the stretching constant, l_i is the actual distance between beads and l_0 is the optimum bead–bead distance. Based on (25,26), the value of K_s was set at 100 $k_B T/l_0^2$ (as it reproduces the correct DNA bond variance), with k_B the Boltzmann constant and T the temperature. The bending energy is determined as:

$$E_B = 0.5 K_B \sum_{i=1}^{N-2} \beta_i^2 \quad (3)$$

where K_B is the bending constant and β_i is the Euler bending angle between the local coordinate systems

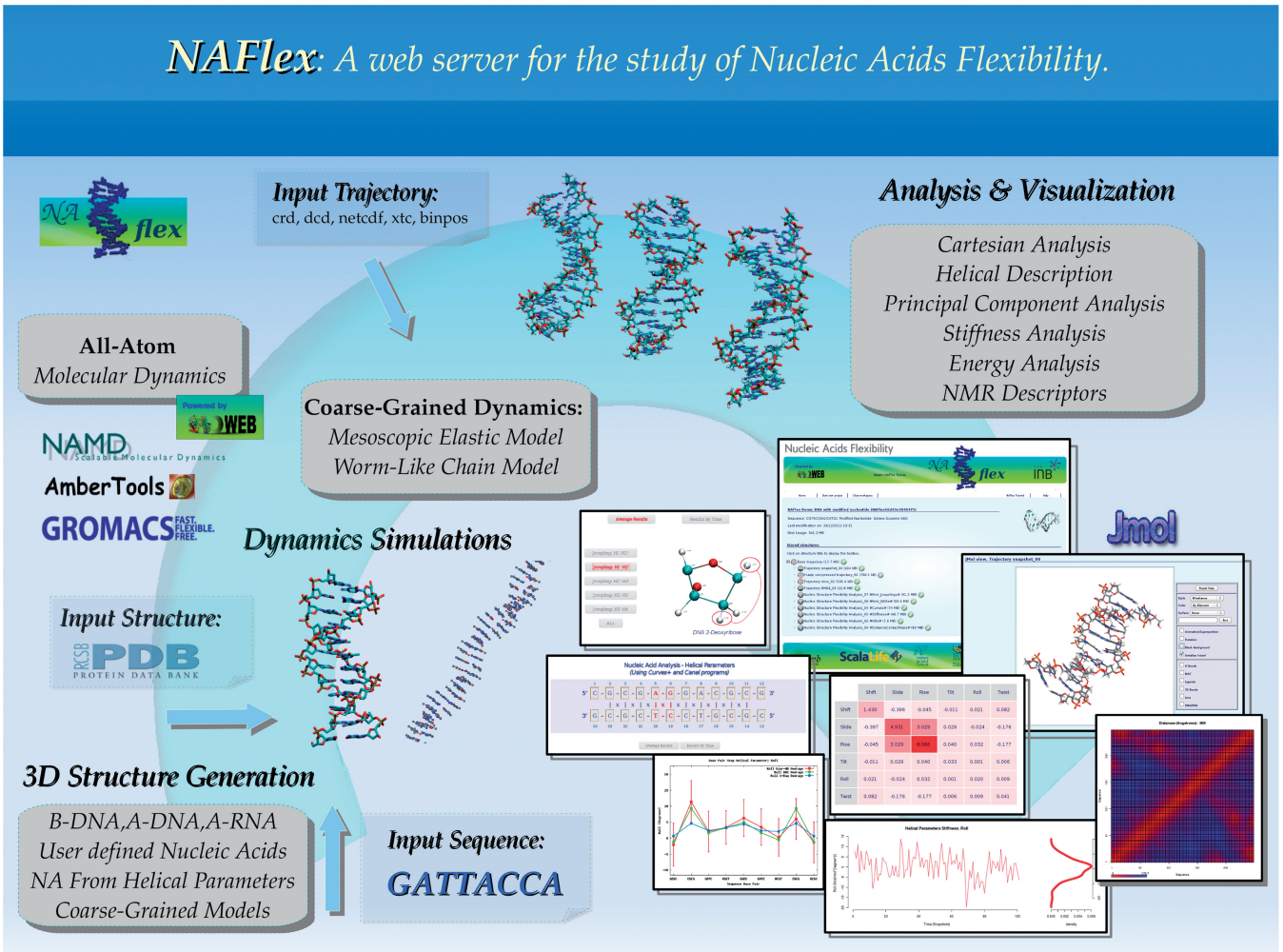

Figure 1. Representation of the NAFlex web server overview showing its three main blocks: (**i**) the inputs accepted: Sequence, structure and trajectory; (**ii**) the simulation engines implemented: molecular dynamics, mesoscopic simulations and WLC calculations; and (**iii**) the set of nucleic acid-specific flexibility analysis and visualization tools offered.

of two consecutive beads; and the torsion potential is defined by:

$$E_T = 0.5K_T \sum_{i=1}^{N-1} (\alpha_i + \gamma_i - \phi_0)^2 \quad (4)$$

where K_T stands for the torsional rigidity, and the sum of α and γ Euler angles defines the torsion between the local coordinate systems of two consecutive beads. ϕ is a parameter that gives the mean DNA twist, and it depends on the bead–bead equilibrium distance and the helical repeat. The values of the bending and torsional rigidity constants are related to the DNA bending (P) and twist (C) persistence lengths, respectively, and the bead–bead equilibrium distance:

$$K_B = \frac{Pk_BT}{l_0} \quad (5)$$

and

$$K_T = \frac{Ck_BT}{l_0} \quad (25,26) \quad (6)$$

The electrostatic repulsion energy (E_{ele}), the only nonlocal term in the Hamiltonian, is determined using Debye–Hückel potential:

$$E_{ele} = \frac{\nu^2 l_0^2}{D} \sum_{i,j} \frac{e^{-\kappa r_{ij}}}{r_{ij}} \quad (7)$$

where ν is the salt-dependent Stigter's effective DNA linear charge density (27), D is the dielectric constant of water, κ is the inverse Debye length and r_{ij} is the distance between two beads. Alternatively, the user can define a value for the effective DNA-bead charge, q_{DNA}, and account for the electrostatic potential as follows:

$$E_{DH} = \frac{qDNA^2}{4\pi\varepsilon\varepsilon_0} \sum_i \sum_{j \neq i} \frac{e^{-Kr_{ij}}}{r_{ij}} \quad (8)$$

where ε_0 is the electric permittivity of vacuum, and ε is the dielectric constant (set to 80).

The elastic mesoscopic model provides an intermediate level of resolution and potential energy complexity (7,8,11,12,14–16). It assumes that DNA deformations can be approximated as the addition of harmonic

distortions of equilibrium base-pair step geometries. Three rotational (twist, roll and tilt) and three translational (slide, shift and rise) degrees of freedom are considered, thereby allowing us to define the Hamiltonian as:

$$E = \Xi\,(\Delta X)^2$$

$$with\ \Xi = \begin{pmatrix} k_w & k_{wr} & k_{wt} & k_{ws} & k_{wl} & k_{wf} \\ k_{wr} & k_r & k_{rt} & k_{rs} & k_{rl} & k_{rf} \\ k_{wt} & k_{rt} & k_t & k_{st} & k_{tl} & k_{tf} \\ k_{ws} & k_{rs} & k_{st} & k_s & k_{ls} & k_{lf} \\ k_{wl} & k_{rl} & k_{tl} & k_{ls} & k_l & k_{lf} \\ k_{wf} & k_{rf} & k_{tf} & k_{lf} & k_{lf} & k_f \end{pmatrix} = k_B T \subset^{-1} \quad (9)$$

where k_B is the Boltzman constant, T is the absolute temperature, E is the energy associated with the deformation ΔX and k_{XY} stands for the different stiffness constants defined by the 36 elements of the stiffness matrix (Ξ) [twist (w), roll (r), tilt (t), rise (s), slide (l) and shift (f)]. The Ξ can be calculated (see equation 9) by inversion of the covariance matrix obtained from either analysis of MD trajectories (at dinucleotide or tetranucleotide level, 22–24) or from the analysis of dinucleotide step variability in the crystal structures of DNAs and DNA–protein complexes (1,15). In all the cases, sampling is obtained using Monte Carlo simulations in the helical coordinate space following the DNALive (28) protocol.

Atomistic molecular dynamics (MD) is the most flexible, universal and accurate simulation engine incorporated to NAFlex. Unfortunately, it is also the most expensive, thus limiting its practical use to medium-sized nucleic acid structures (7–10). After the quality control check, and before running a MD simulation, the system is neutralized, immersed in the desired solution, minimized, thermalized and equilibrated in a few steps (Supplementary Figure S1). Once the equilibration is finished, the server provides NAMD (29), AMBER (30) or GROMACS (31) adapted input files, which can be then used to perform the production run on the user's local systems (see server help). Output trajectories can be uploaded back to the server for analysis.

Analysis tools

NAFlex integrates a variety of analysis packages for mining nucleic acid trajectories. Analysis results are presented in several numerical and graphical formats. Available analyses include the following: standard Cartesian and helical analysis (17,32); principal component analysis (PCA) (33,34) to determine the essential deformation movements and their associated stiffness; and helical stiffness analysis, performed following Olson–Lankas's approach (14,15). A basic energy analysis is conducted to determine stacking and hydrogen bonding interaction energies along the trajectory. Finally, NAFlex also provides estimates of a number of NMR observables (see help for details), which can be used for either trajectory validation or for *ab initio* spectral predictions.

SERVER IMPLEMENTATION AND USAGE

NAFlex is a web portal implemented in PHP and MySQL. It provides users with a personal workspace where intermediate data, trajectories and results of analysis can be stored. Access to the server is free; however, registration is required to maintain a permanent workspace. Nonregistered users can use the full functionality of the server, although they need to download the results at the end of the session. In the present implementation, registered users are provided with 2 Gb of storage space. Storage for specific projects can be increased on demand. NAFlex is powered by the MDWeb platform, which was designed for automated MD simulations of proteins (17). MDWeb provides the necessary modules for data management, structure checking, MD simulation setup and standard Cartesian trajectory analysis (see http://mmb.irbbarcelona.org/MDWeb/). NaFlex incorporates a variety of tools specifically oriented for nucleic acids, including input facilities, simulation engines and analysis tools (see 'Materials and Methods' section).

Workspace

The starting points for the analysis could be single structures, uploaded as PDB files or derived from a given sequence, or previously simulated trajectories (see above). The input structure or trajectory is used as initial step to initiate a project in the personal workspace. There is no limit to the number of initiated projects while the permitted storage capacity is not exceeded. For each project, the workspace holds all intermediate structures and results, organized as a tree view (Supplementary Figure S2), thereby allowing the user to track the history of operations performed. For each entry of the tree, a number of operations, selected on the basis of the type of data, are offered to the user. For all intermediate results, a download tool and a series of JMol-powered (http://www.jmol.org) visualization tools are available. Complete projects can also be downloaded for local storage. Downloaded projects can be restored in NAFlex again at a later time, thus recovering the original workspace.

Simulations

Coarse-grained simulations from initial structures can be performed within the NAFlex environment. When atomistic MD simulations are done, a series of preparation tools is required. Specific setup procedures adapted to nucleic acids are available, although experienced users may design and incorporate their own protocols into the workflow. Running short MD simulations for testing purposes is allowed on the server; however, production simulations should be run locally in the user's own facilities using NAFlex-generated input files (see above). The user can then upload trajectories into his/her workspace for analysis.

Analysis tools

NAFlex provides a wide repertoire of specific tools for the analysis of nucleic acids, some of them local, others

incorporated from external sources. All the tools are offered in a common interface, which requires no additional expertise for its use. The user has complete control over the level of resolution in the analysis through an interactive duplex viewer (Figure 2a). He/she can also analyze time course or average values and study in detail the resulting structural models (see Figure 2b and c). When available, standard values obtained from the literature are included for comparison purposes (Figure 2b). Plots can be opened in separate windows to compare the structural or mechanical behavior of different regions of the molecules. In all cases, raw data can be downloaded for local analysis. It should be noted that full analyses are done upon the initial request. This allows the user to browse the results even for a large structure without noticeable delay. Trajectories can be manipulated to extract individual snapshots for further inspection (Figure 2d).

Please note that trajectory upload is limited by network bandwidth. The current limit for NAFlex-uploaded trajectories is 100 Mb (as indicated in the corresponding help pages). However, in our experience, once the trajectory has been stripped of solvent and ions, and taking a representative ensemble of snapshots, the analysis of the resulting trajectory with NAFlex can provide useful information about the general flexibility of the molecule. See Example 4 in the next section as an illustration of a real analysis.

EXAMPLES OF USE

NAFlex is an extremely flexible server that offers a large number of options, and it does not require deep knowledge of either simulation engines or the physics of nucleic acids. A few examples of use are included here as references. These examples can be also accessed online on the NAFlex server using the 'demo' account.

Example 1: Atomistic MD from a nucleotide sequence

The files required to run an atomistic MD simulation can be obtained following a robust protocol. First of all, the user should define the type of input (Supplementary Figure S3a), in this case, *DNA/RNA Simulation From sequence*. The only inputs required are the nucleotide sequence (either typed or from a FASTA-formatted file) and the desired set of DNA/RNA helical parameters (Supplementary Figure S3b). For instance, to generate a DNA molecule with a higher value for twist than the canonical B-DNA, *User-defined DNA* should be chosen at *DNA/RNA type* selector and the *Twist* value should be modified accordingly (in the example from 36.0° to 38.0°). The structure will then be generated using the Nab program from the Ambertools package (19). A new window will appear showing the newly created structure in a JMol applet, together with the results of a set of structural checks (Supplementary Figure S3c and d). After validation of the original structure, a new workspace for the project is prepared (see Supplementary Figure S3e). The interface allows the user to select the required operation, in this case a complete setup for an AMBER MD simulation (see Supplementary Figure S3e) using the ff99SB (35) and PARMBSC0 (36) force fields with recent corrections [ildn (37) and OL3 (38)]. These are the recommended settings for a wide set of structures, including protein–DNA complexes and RNA; however, users have the choice of other force-field combinations. The evolution of the running workflow can be visualized in real time (Supplementary Figure S3f), and the final files can be downloaded for local production runs (Supplementary Figure S3g).

Example 2: Mesoscopic simulation of DNA

This example illustrates a mesoscopic simulation of a long protein-free DNA molecule at the base-pair level of resolution. As in the previous example, the first steps are the definition of the type of input (*DNA/RNA Simulation From sequence*), the selection of a title for the project and the introduction of the sequence (see Supplementary Figure S4a and b). In this case, *DNA/RNA type* should be set to *Coarse-Grained DNA Model (Base Step Level)*, to select the desired coarse-grained resolution level. The prepared structure can be visualized using the JMol applet (Supplementary Figure S4c). To launch the simulation, the user should select the desired number of ensemble snapshots (500 in this example), while the type of mesoscopic Hamiltonian (*Coarse-grained DNA Elastic Mesoscopic Model*) is already defined by the resolution level (see Supplementary Figure S4d). After the Monte Carlo ensemble has been collected, the snapshots can be further analyzed to trace the flexibility of the fiber being studied. In this particular example, the probability of cross-talk between distant segments of the DNA is analyzed. This can be done by simple inspection of the ensemble (Supplementary Figure S4e) and from the averaged nucleotide-contact maps (see Supplementary Figure S4f). The appearance of short contacts (red) off-diagonal is indicative of long-range contacts of potential biological relevance.

Example 3: Nucleic acid flexibility analysis

The example illustrates the analysis of a 100-ns MD simulation of a DNA dodecamer (dCGCGAGGACGCG). The trajectory in this case was uploaded from an external source. Analysis is initiated by selecting the *Nucleic Acid Flexibility Analysis* operation in the *Analysis Tools* box (Supplementary Figure S5a). The first results to be inspected are the helical characteristics of the duplex, obtained from a CURVES (32) analysis. Results (Supplementary Figure S5b and c) show the presence of a subtle *Twist* bimodality (Time-course *Twist* plot, Supplementary Figure S5b) and a marked *Roll* increase (taking X-ray as reference) in the first CpG step (Averaged *Roll* plot, Supplementary Figure S5c). This observation suggests that this step can be especially flexible and can display spontaneous curvature. Additional analysis shows that the DNA is highly robust in terms of helical structure, with strong hydrogen bonds and stacking interactions (see Supplementary Figure S5d); however, both ends show fraying effects (see, Supplementary Figure S5e), which are expected also to impact Nuclear Overhauser Effects (NOEs) (see Supplementary Figure S5f).

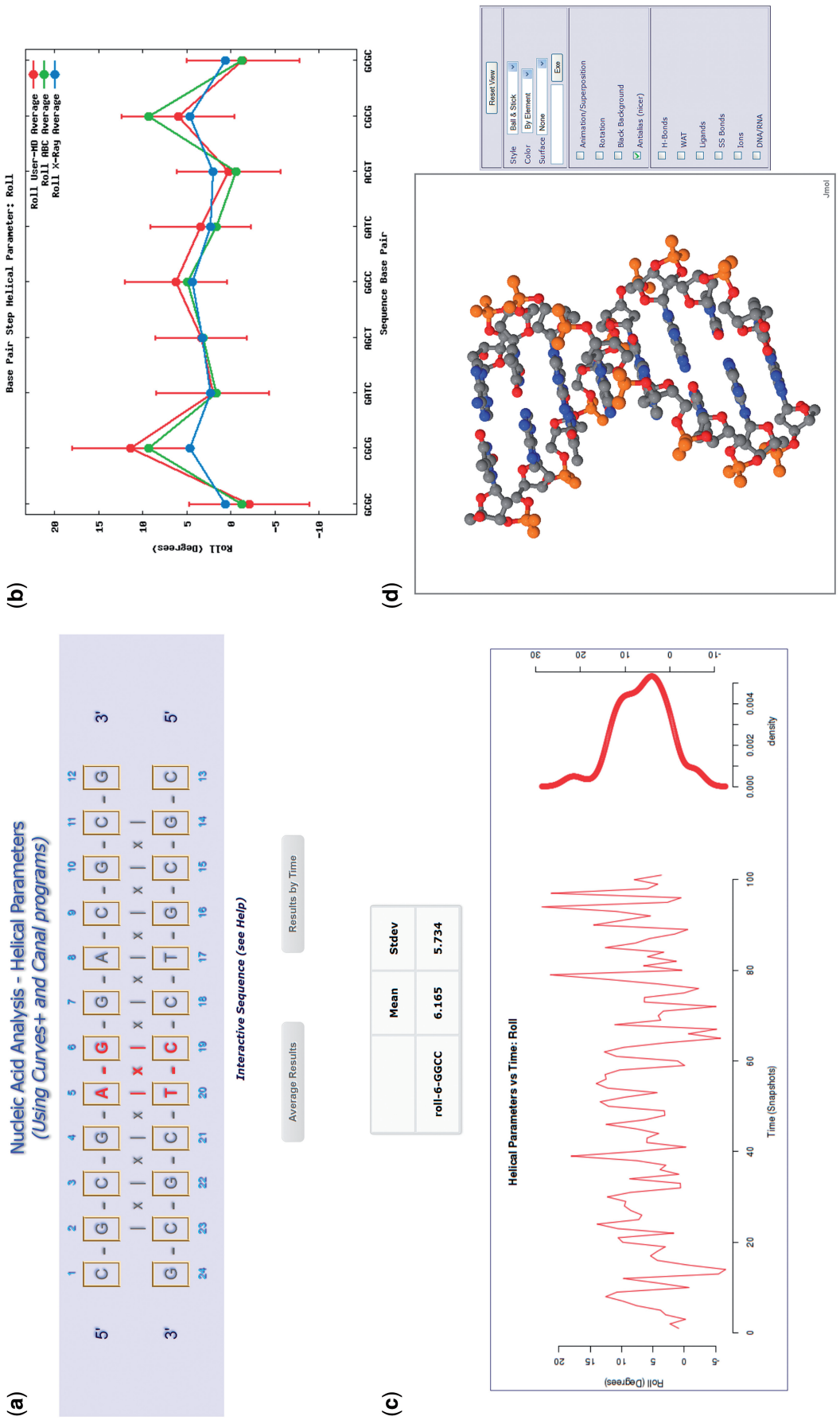

Figure 2. Screenshots of the analysis results interface. (**a**) Interactive duplex viewer allowing the level of resolution where the analysis is focussed: base-pairs *(|)*, base-pair steps *(X)* and nucleotide pair steps *(-)*. (**b**) Sequence distribution of time-averaged *Roll*, compared with MD- and X-ray averaged values. (**c**) Time course of the *Roll* at *6-GGCC* base-pair step, with an accessory histogram (mean and standard deviations are also displayed). (**d**) JMol interactive representation of a nucleic acid structure. Graphical visualization is offered for structures and also for trajectories.

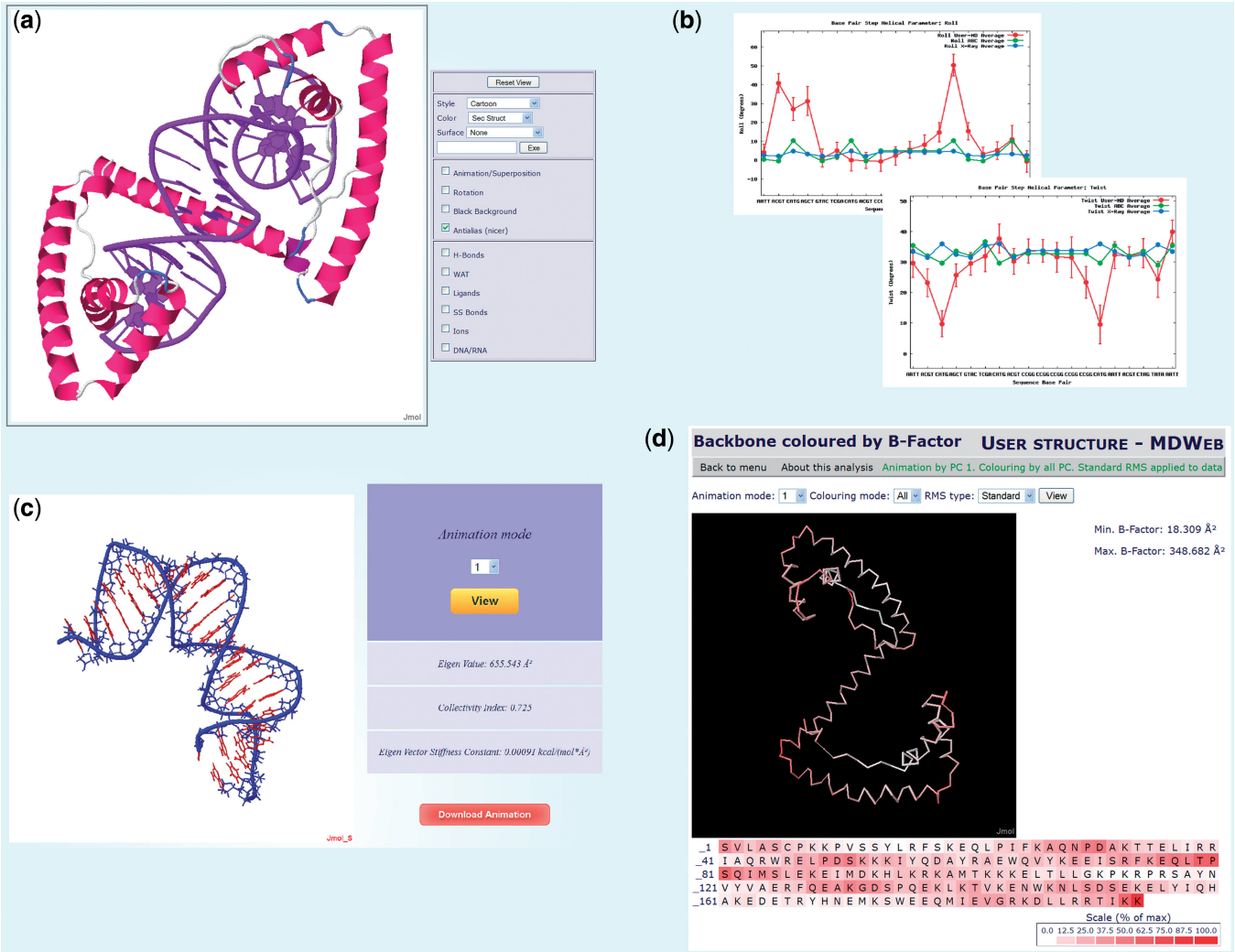

Figure 3. Protein–DNA complex TFAM–LSP as a case study. (**a**) JMol representation of the TFAM–LSP complex. To facilitate inspection, the protein part is represented by a dark pink cartoon, while the nucleic part is shown by a purple cartoon. (**b**) Average inter-base-pair helical parameters *Roll* and *Twist* computed on the LSP oligonucleotide plotted together with standard values obtained from the literature for comparison purposes. (**c**) Snapshot of the animation of the first mode of the DNA obtained from PCA. (**d**) B-factor analysis of the TFAM protein.

Example 4: Protein–DNA complex, human mitochondrial transcription factor A as a case study

The example shows the analysis of an important protein–DNA complex. The structural experimental information available (39–41) suggests that flexibility is crucial for the functionality of the complex. Analysis is initiated from a user-provided 100-ns trajectory obtained with GROMACS 4 (31). In the example, a 1-ns time window is retained for analysis.

A JMol visualizer provides initial structural views of the trajectory. Thus, in Figure 3a, the human mitochondrial transcription factor A (TFAM) protein can be clearly identified in dark pink cartoon representation, with the two high mobility group (HMG) protein domains joined by the inter-domain linker helix. The 22-base-pair oligonucleotide mitochondrial light strand promoter (LSP) attached to the TFAM protein is shown in purple. Once uploaded, the trajectory is automatically split into protein and DNA, allowing separate analysis. An extensive flexibility analysis of the protein counterpart (TFAM) can be performed using FlexServ platform (42), also available through the NAFlex workspace.

Basic helical analysis (see Figure 3b) shows that *roll* and *twist* angles are completely distorted along the trajectory in the regions where the protein residues intercalate into DNA base-pairs (Figure 3a). PCA was used to obtain a clear picture of the essential movements of the oligonucleotide in the complex. In this example, the first three eigenvectors explain most of the variance (Figure 3c and interactive JMol online). The first two modes involve the kinking of the complete LSP along the direction of the linker helix, bringing both DNA ends closer together, thus adopting a U-turn. The third mode shows the twisting of the nucleic acid caused by the motion of the two TFAM domain boxes HMG1 and HMG2. Domain motions can be visualized using the available JMol Applet. In accordance with the hypothesis that the HMG domains induce an overall DNA U-turn stabilized by the helix linker (40,41), FlexServ analysis of the protein moiety (Figure 3d) reveals the presence of flexible

segments near the domain boxes, while the linker helix appears to be relatively rigid.

CONCLUSIONS

The field of nucleic acid simulation has reached maturity. It has moved away from the times when only 'proof of concept' simulations were done by a very small number of highly specialized groups toward a 'full production' situation, where nucleic acid simulation tools are used by many groups, often not experts in the physics of nucleic acids or the theory behind the simulation package. NAFlex is a bioinformatics tool created to facilitate the use of simulation tools for nonexpert users interested in gaining insight into the dynamics of nucleic acid systems.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–5.

ACKNOWLEDGEMENTS

We thank Drs G. Portella, F. Battistini, A. Alibés and P. Dans for their extensive testing of the platform, P. Sfriso for his help in NOE intensity computation, P. Andrio for his help in the design of the graphical interface and J. Alcántara for the design and maintenance of the computers used by the server. We are also indebted to the developers of the CURVES+, CANAL, GROMACS, AMBERTOOLS, NAMD, VMD and JMol software, which are being used in the server.

FUNDING

Funding for open access charge: *Spanish Ministerio de Economía y Competitividad* [BIO2012-3286]; European Research Council Advanced Grant (ERC); *Instituto Nacional de Bioinformática* (INB); Consolider E-Science Project; EU-Scalalife project; Fundación Marcelino Botín; European Union's Seventh Framework Programme (FP7/2007–2013) [275096 to R.C.-G. and M.O.].

Conflict of interest statement. None declared.

REFERENCES

1. Dans,P., Pérez,A., Faustino,I., Lavery,R. and Orozco,M. (2012) Exploring polymorphisms in B-DNA helical conformations. *Nucleic Acids Res.*, **40**, 10668–10678.
2. Pérez,A., Noy,A., Lankas,F., Luque,F.J. and Orozco,M. (2004) The relative flexibility of DNA and RNA: database analysis. *Nucleic Acids Res.*, **32**, 6144–6151.
3. Zheng,G., Colasanti,A.V., Lu,X.J. and Olson,W.K. (2010) 3DNA Landscapes: a database for exploring the conformational features of DNA. *Nucleic Acids Res.*, **38**, D267–D274.
4. Nikolova,E.N., Bascom,G.D., Andricioaei,I. and Al-Hashimi,H.M. (2012) Probing sequence-specific DNA flexibility in A-tracts and pyrimidine-purine steps by nuclear magnetic resonance 13C relaxation and molecular dynamics simulations. *Biochemistry*, **51**, 8654–8664.
5. Bothe,J.R., Lowenhaupt,K. and Al-Hashimi,H.M. (2011) Sequence-specific B-DNA flexibility modulates Z-DNA formation. *J. Am. Chem. Soc.*, **133**, 2016–2018.
6. Nikolova,E.N., Kim,E., Wise,A.A., O'Brien,P.J., Andricioaei,I. and Al-Hashimi,H.M. (2011) Transient Hoogsteen base pairs in canonical duplex DNA. *Nature*, **470**, 498–502.
7. 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.
8. Orozco,M., Noy,A. and Pérez,A. (2008) Recent advances in the study of nucleic acids flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.*, **18**, 185–193.
9. Pérez,A., Luque,F.J. and Orozco,M. (2012) Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.*, **45**, 196–205.
10. Laughton,C.A. and Harris,S.A. (2011) The atomistic simulation of DNA. *WIREs Comput. Mol. Sci.*, **1**, 590–600.
11. Drsata,T. and Lankas,F. (2013) Theoretical models of DNA flexibility. *WIREs Comput. Mol. Sci.*, 14 Feb (doi:10.1002/wcms.1144; epub ahead of print).
12. Lankas,F. (2012) Modelling nucleic acid structure and flexibility: from atomic to mesoscopic scale. In: Schlick,T. (ed.), *Innovations in Biomolecular Modeling and Simulations*, Vol. 2. Royal Society of Chemistry, London, pp. 3–32.
13. Allison,S.A. (1986) Brownian dynamics simulation of wormlike chains. Fluorescence depolarization and depolarized light scattering. *Macromolecules*, **19**, 118.
14. Lankas,F., Sponer,J. and Langowski,J. (2000) Sequence-dependent elastic properties of DNA. *J. Mol. Biol.*, **299**, 695–709.
15. 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. USA*, **95**, 11163–11168.
16. Olson,W.K. (1996) Simulating DNA at low resolution. *Curr. Opin. Struct. Biol.*, **6**, 242–256.
17. Hospital,A., Andrio,P., Fenollosa,C., Cicin-Sain,D., Orozco,M. and Gelpí,J.L. (2012) MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations. *Bioinformatics*, **28**, 1278–1279.
18. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
19. Case,D.A., Darden,T.A., Cheatham,T. III, Simmerling,C.L., Wang,J., Duke,R.E., Luo,R., Walker,R.C., Zhang,W., Merz,K.M. *et al.* (2012), AMBER 12. San Francisco: University of California.
20. Arnott,S., Hukins,D.W.L., Dover,S.D., Fuller,W. and Hodgson,A.R. (1973) Structures of synthetic polynucleotides in the A-RNA and A'-RNA conformations. X-ray diffraction analyses of the molecular conformations of polyadenylic acid—polyuridylic acid and polyinosinic acid—polycytidylic acid. *J. Mol. Biol.*, **81**, 107–122.
21. Lakshiminarayanan,A.V. and Sasisekharan,V. (1970) Stereochemistry of nucleic acids and polynucleotides II. Allowed conformations of the monomer unit for different ribose puckerings. *Biochim. Biophys. Acta*, **204**, 49–59.
22. Pérez,A., Lankas,F., Luque,F.J. and Orozco,M. (2008) Towards a consensus view of B-DNA flexibility. *Nucleic Acids Res.*, **36**, 2379–2394.
23. Faustino,I., Pérez,A. and Orozco,M. (2010) Towards a Consensus view of duplex RNA flexibility? *Biophys. J.*, **99**, 1876–1885.
24. Lavery,R., Zakrzewska,K., Beveridge,D., Bishop,T.C., Case,D., Cheatham,T. III, Dixit,S., Jayaram,B., Lankas,F., Laughton,C.A. *et al.* (2010) A systematic molecular dynamics study of the nearest-neighbor effects on base pair and base step conformations and fluctuations in B-DNA. *Nucleic Acids Res.*, **38**, 299–311.
25. Jian,H., Vologodskii,A.V. and Schlick,T. (1997) A combined wormlike-chain and bead model for dynamic simulations of long linear DNA. *J. Comp. Phys.*, **136**, 168–179.
26. Jian,H., Schlick,T. and Vologodskii,A.V. (1998) Internal motions of supercoiled DNA: brownian dynamics simulation of site juxtaposition. *J. Mol. Biol.*, **284**, 287–296.
27. Stigter,D. (1977) Interactions of highly charged colloidal cylinders with applications to double-stranded DNA. *Biopolymers*, **16**, 1435–1448.

28. Goñi,J.R., Fenollosa,C., Pérez,A., Torrents,D. and Orozco,M. (2008) DNAlive: a tool for the physical analysis of DNA at the genomic scale. *Bioinformatics*, **24**, 1731–1732.
29. Phillips,J.C., Braun,R., Wang,W., Gumbart,J., Tajkhorshid,E., Villa,E., Chipot,C., Skeel,R.D., Kalé,L. and Schulten,K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
30. Case,D.A., Cheatham,T.E. III, Darden,T., Gohlke,H., Luo,R., Merz,K.M. Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
31. Hess,B., Kutzner,C., van der Spoel,D. and Lindahl,E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.
32. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
33. Amadei,A., Linssen,A.B. and Berendsen,H.J. (1993) Essential dynamics of proteins. *Proteins*, **17**, 412–425.
34. Noy,A., Meyer,T., Rueda,M., Ferrer,C., Valencia,A., Pérez,A., de la Cruz,X., López-Bes,J.M., Luque,F.J. and Orozco,M. (2006) Datamining of molecular dynamics trajectories of nucleic acids. *J. Biomol. Struct. Dyn.*, **23**, 447–455.
35. Hornak,V., Abel,R., Okur,A., Strockbine,B., Roitberg,A. and Simmerling,C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, **65**, 712–725.
36. 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 acid simulations. Improving the representation of α/β conformations. *Biophys. J.*, **92**, 3817–3829.
37. Lindorff-Larsen,K., Piana,S., Palmo,K., Maragakis,P., Klepeis,J.L., Dror,R.O. and Shaw,D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950–1958.
38. Zgarbova,M., Otyepka,M., Sponer,J., Mladek,A., Banas,P., Cheatham,T.E. and Jurecka,P.J. (2011) Refinement of the Cornell et al. nucleic acid force field based on reference quantum chemical calculations of torsion profiles of the glycosidic torsions. *J. Chem. Theory Comput.*, **7**, 2886–2902.
39. Kaufman,B.A., Durisic,N., Mativetsky,J.M., Costantino,S., Hancock,M.A., Grutter,P. and Shoubridge,E.A. (2007) The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. *Mol. Biol. Cell.*, **18**, 3225–3236.
40. Rubio-Cosials,A., Sidow,J.F., Jiménez-Menéndez,N., Fernández-Millán,P., Montoya,J., Jacobs,H.T., Coll,M., Bernadó,P. and Solà,M. (2011) Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat. Struct. Mol. Biol.*, **18**, 1281–1289.
41. Rubio-Cosials,A. and Solà,M. (2013) U-turn DNA bending by human mitochondrial transcription factor A. *Curr. Opin. Struct. Biol.*, **23**, 116–124.
42. Camps,J., Carrillo,O., Emperador,A., Orellana,L., Hospital,A., Rueda,M., Cicin-Sain,D., D'Abramo,M., Gelpí,J.L. and Orozco,M. (2009) FlexServ: an integrated tool for the analysis of protein flexibility. *Bioinformatics*, **25**, 1709–1710.