3D-DART: a DNA structure modelling server

Marc van Dijk and Alexandre M. J. J. Bonvin*

Bijvoet Center for Biomolecular Research, Science Faculty, Utrecht University, The Netherlands

ABSTRACT

There is a growing interest in structural studies of DNA by both experimental and computational approaches. Often, 3D-structural models of DNA are required, for instance, to serve as templates for homology modeling, as starting structures for macro-molecular docking or as scaffold for NMR structure calculations. The conformational adaptability of DNA when binding to a protein is often an important factor and at the same time a limitation in such studies. As a response to the demand for 3D-structural models reflecting the intrinsic plasticity of DNA we present the 3D-DART server (*3D*NA-Driven *D*NA *A*nalysis and *R*ebuilding *T*ool). The server provides an easy interface to a powerful collection of tools for the generation of DNA-structural models in custom conformations. The computational engine beyond the server makes use of the 3DNA software suite together with a collection of home-written python scripts. The server is freely available at http://haddock.chem.uu.nl/dna without any login requirement.

INTRODUCTION

DNA often changes its conformation as a result of interactions with various ligands; especially binding to proteins can result in large conformational changes such as helical kinks (1) or local helical untwisting (2). These play an important roll in providing complementarity to the protein binding surface and contributing to the interaction specificity (3). In order to fully understand the nature of the conformational changes taking place upon complex formation, 3D, atomic-resolution structures are required. Experimental methods such as X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR) but also computational approaches such as macro-molecular docking are important techniques for obtaining such 3D structures or models.

Most techniques make use of 3D-structural models of DNA at some point along the structure calculation pipeline. NMR for instance can benefit from the regularity in the structure of double-stranded DNA by using a model as starting point for structure calculations, thereby compensating for the lack of long range structural information. For macro-molecular docking, a starting model is often required as experimental structures might not be available. Often, starting from multiple models with different conformations improves the results. Finally, as last example, homology-modeling programs require a template model as starting point for the homology building process.

The regularity in the structure of double-stranded DNA makes it especially suitable for modeling. Various software packages are available that convert a user specified base-pair sequence into a 3D structure using regular nucleotide building blocks (4–8). However, most of these software packages, some of which are available via webservers (4,7), are only able to generate models in ideal canonical conformations (5,7) or in conformations mimicking that of a free unbound structure (4).

The structures of double-stranded DNA in complex with various ligands often show considerable conformational changes compared to their unbound counterparts (9–12). This plasticity originates at a 'local' level in the orientation of one base relative to its Watson–Crick partner and of two base pairs relative to one another. These 'local' changes accumulate and result in bending and twisting of the structure at a 'global' level. Only a few existing programs, such as NAMOT (8) and NAB (6), offer options to introduce custom bends in the generated DNA conformation and give control over all local parameters; they however require some expertise from the user and are not available as web servers. Here we describe the 3D-DART web server (*3D*NA-Driven *D*NA *A*nalysis and *R*ebuilding *T*ool) which we developed to allow for the easy generation of 3D-structural DNA models with a defined conformation by providing control over both 'global' and 'local' conformational features.

The generation of models is accomplished by modification of the well-established rotational and translational parameters that describe the position of one base to its Watson–Crick counterpart and of two successive base pairs relative to one another (13). It has been demonstrated in the past that rebuilding a double-stranded DNA structure using these parameters results in a near native structure (5). The only exceptions are local changes in the sugar and phosphate backbone conformation.

*To whom correspondence may be addressed. Tel: +31 30 2533859; Fax: +31 30 2537623; Email: a.m.j.j.bonvin@uu.nl

The 3D-DART server uses the Roll, Tilt and Twist parameters to introduce bends into the structure. The other parameters can be used to 'fine-tune' the conformation of the structure. Note that 3D-DART does not provide custom control of the sugar-phosphate backbone conformation (in contrast for example to NAMOT). The 3DNA software (5) is used to generate a 3D-structural model from the modified parameters.

The server accepts a nucleotide sequence, a base pair (step) parameter file or a DNA-containing PDB coordinate file as input. The server returns 3D-structural models with the desired conformation as well as a collection of analysis and intermediate files. Several additional and convenient functions are available to control the markup of the resulting PDB coordinate files, for instance to prepare them for use in the macro-molecular docking program HADDOCK (14,15) also developed in our group. For the same purpose the server can automatically generate a DNA restraint file (16) as an additional feature. The server is freely available at http://haddock.chem.uu.nl/dna without any login requirement.

3D-DART MODELING PROCEDURE

Local bending is often at the origin of double-stranded DNA distortions when in complex with various proteins (9–11). This type of bending can be described in terms of the vector between two successive base pairs (a base pair step). The length of this vector (Figure 1, thick black arrows) describes the distance between the two base pairs in a base-pair step, also known as Rise. It usually does not vary much. In unbent canonical DNA, these vectors align with the Z-axis that represents the main helical path of the structure (Figure 1A). When the DNA is bent, then the position of the vector relative to the global reference frame describes the magnitude and orientation of the bend angle. The magnitude of the bend corresponds to the vector component projected on the Y–Z plane and its orientation to the component projected on the Y–X plane (Figure 1B). The accumulation of successive vectors then determines the overall bend in the structure.

Such a bend vector can be decomposed into a Roll, Tilt and Twist base pair step parameter contributions. The 3D-DART server uses these parameters to introduce bends in the structure. The underlying algorithm is based on the transformation of the global bend vector in Euclidean space into Roll and Tilt values in the local base pair step reference frame (Figure 1C). This transformation is accomplished using the following steps:

(i) The definition of bend vectors requires the definition of an origin in Euclidean space set to an arbitrary base pair in the sequence (ri) such that the main helical path of the structure is aligned with the Z-axis. By default the central base pair is chosen (Figure 1A, cyan).

(ii) Because the helical DNA structure is twisted the orientation component of the bend vector (O_i, $0° \rightarrow 360°$) at a given base-pair step i needs to be corrected for the local Twist value. In Figure 1A
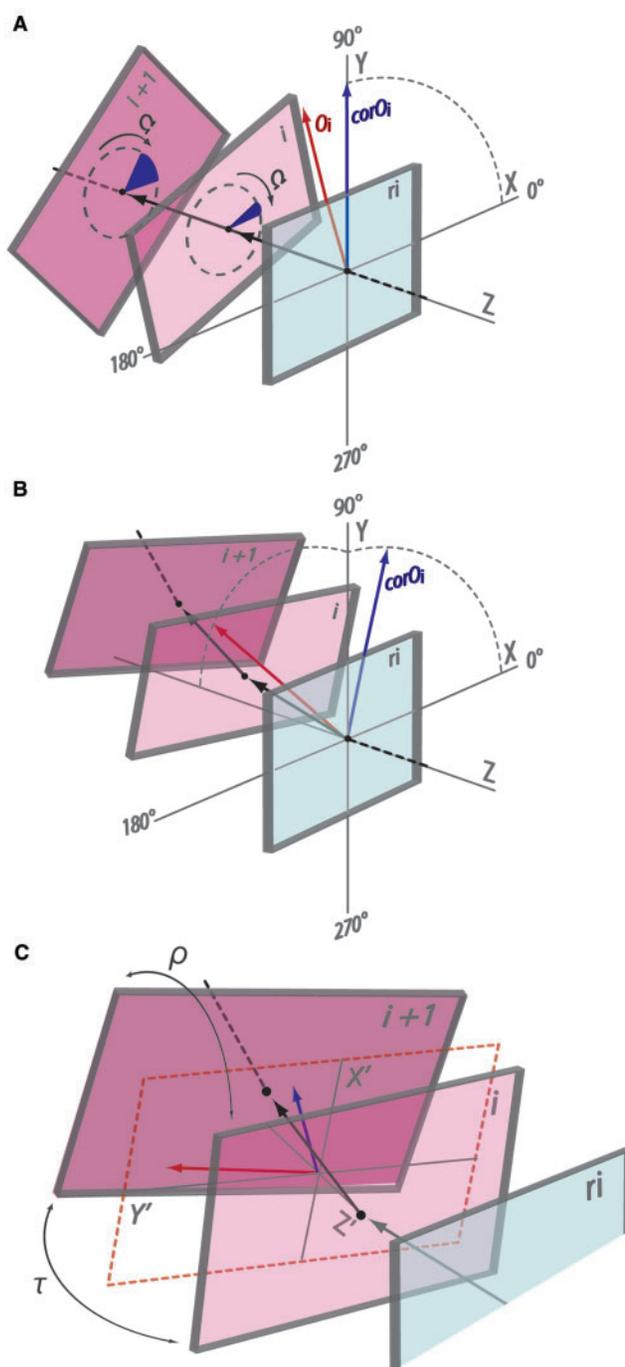

Figure 1. One block per base-pair Calladin–Drew plot of DNA illustrating the relation between the local Twist (Ω), Roll (ρ) and Tilt (τ) values and the global bend angle for a given base-pair step. Vector projections are normalized for illustrative purposes. (**A**) The correction of the orientation component of the bend vector for the local Twist value at base pair i and $i + 1$ (blue circle parts). The red arrow indicates the value of the orientation component (O_i) before Twist correction and the blue arrow ($corO_i$ aligned with Y-axis) after correction. (**B**) A bend in the structure as a result of a different bend angle vector (thick black arrows) between every successive base-pair step. The blue arrow illustrates the orientation component of the vector (Y–X plane) and the red arrow the magnitude (Y–Z plane). (**C**) provides a detailed view of the local base-pair step reference frame between base i and $i + 1$. The global bend vector (thick black arrow) is decomposed into a Tilt (red arrow, Y'–Z' plane) and Roll (blue arrow, X'–Z' plane) contribution.

the red arrow illustrates the direction of the orientation component before Twist (Ω) correction and the blue one after correction. In Equation (1) the orientation component O_i is corrected for the Twist value at base pair step i by subtracting the accumulated Twist from ri to i (Figure 1A, blue circle partitions). The accumulated Twist value is positive above and negative below the reference base pair.

$$corO_i = O_i + \sum_{ri=1}^{i} \Omega_i \qquad 1$$

(iii) The bend vector with the corrected orientation component ($corO_i$) at base pair step i is the result of the Roll (ρ) and Tilt (τ) in the local base pair step reference frame of i (Figure 1C). The base pair tilt caused by Roll ($X'–Z'$ plane) is orthogonal to the base pair tilt originating from the Tilt parameter ($Y'–Z'$ plane). Together they can span $360°$. Equation (2) defines the fractional contributions of the Roll ($fr\rho_i$) and Tilt ($fr\tau_i$) to the vectors orientation:

$$fr\tau_i = \cos(corO_i \times \pi/180)$$
$$fr\rho_i = \sin(corO_i \times \pi/180) \qquad 2$$

The in Figure 1C blue arrow represents the tilt fraction and the red arrow the roll fraction.

(iv) Finally the actual value of the Roll and Tilt parameters at base pair i are calculated (Equation 3) to reflect both the magnitude and orientation components of the vector representing global bend angle A_i.

$$\rho = \left(\sqrt{A_i^2 \times |fr\rho_i|} \right) \times S_i$$
$$\tau = \left(\sqrt{A_i^2 \times |fr\tau_i|} \right) \times S_i \qquad 3$$

where:

$S_i = 1$ if $fr\rho_i$ or $fr\tau_i > 0$; $S_i = -1$ if $fr\rho_i$ or $fr\tau_i < 0$; $S_i = 0$ if $fr\rho_i$ and $fr\tau_i = 0$

Using this algorithm the magnitude and orientation of a bend angle can be calculated for every individual base pair step in the sequence of a double-stranded DNA structure. The algorithm is based on strict geometrical principles and the resulting structures are not energy minimized. Therefore it is important to emphasize that the conformation does not need to reflect an energetically favorable state. Figure 2 illustrates the procedure with a few examples. Introducing a $3°$ bend with the same orientation in a 20 base pair structure create a smooth bend of $60°$ (Figure 2A). Restricting a $30°$ bend to only the central 3 base pair steps in the same sequence creates a kink (Figure 2B). Using different angle values for every base pair step in the sequence gives precise control over the DNA conformation (Figure 2C).

3D-DART WEB SERVER

Input

The generation of custom DNA 3D-structural models is based on the manipulation of the 6 bp and 6-bp step parameters describing the conformation of the structure. The first step in using the server is the definition of a source for the base-pair (step) parameter template file that serves as a starting point for the modeling process. For this three different options are supported:

- The first option consists of inputting a nucleotide sequence defined as the bases belonging to the $5'–3'$ template strand. The server then uses the 'fiber' module of 3DNA (5) to generate a canonical A- or B-DNA structure with the defined sequence and the 'find_pair' and 'analyze' module of the same software to generate the base pair (step) parameter file.
- The second input option consists of uploading an user-defined PDB coordinate file. The DNA in this file is analyzed and a base pair (step) parameter file representing its conformation is generated. This allows for the introduction of custom changes in an already existing structure.
- Finally, as third input option, the server also accepts a predefined base pair (step) parameter file in 3DNA format.

Parameters

The base pair (step) parameter file that results from the input data is subsequently used to start the modeling phase. At this stage, the user can introduce bends into the structure. There are two modes in which this can be accomplished referred to as 'Global' and 'Local':

- In the *'Global' mode* the defined bend angle is evenly distributed over all base pair steps in the user-defined zone (Figure 2A and B). The 'Global' modeling mode accepts ranges of parameters so that multiple models within a given bend angle and/or bend angle orientation range can be generated. Six models from $10°$ to $40°$ with steps of $5°$ for example.
- In the *'Local' mode* the bend angle and its orientation in Euclidean space can be defined uniquely for every base pair step in the user-defined zone (Figure 2C).

Next to the introduction of bends the user can define custom values for the various base pair and base pair step parameters. These values are subsequently used for every base pair or base pair step in the sequence. If a bend is introduced then of course the Roll and Tilt values will be substituted by the ones needed to introduce the bend. If location-specific values for the parameters need to be introduced it is advised to start from a base pair step parameter file containing these values or from a PDB coordinate file reflecting these values.

The combination of the possible input sources together with the precise modeling options makes the 3D-DART server very versatile, allowing the generation of both ideal DNA models as well as fully customized ones by

Figure 2. Two block per base Calladine–Drew plots of a 20-nt B-DNA structure. The black dotted line defines the main helical path. Two examples of bending using the 'Global' mode of the server are shown: (**A**) A DNA conformation with a smooth bend of 60° distributed evenly over the entire structure and (**B**) a conformation with a 30° kink in which the bend was restricted between base pairs 10 and 12. (**C**) A DNA conformation with custom angles values for every base pair step generated using the 'Local' mode of the server. The asterisk indicates the reference base pair used in the algorithm to generate the bend. The pictures were created using 3DNA (5) and PyMol (DeLano Scientific LLC, Palo Alto, CA, USA).

introducing local conformational changes in already existing structures.

Output

The collection of base pair (step) parameter files resulting from the modeling phase are converted into a 3D structure in PDB format by the 'rebuild' module of 3DNA. The 3D structure is built based on the nucleotide coordinates for the common bases as determined by Arnott *et al.* (17). The generated 3D-structural models are returned in a zipped archive containing in addition a collection of analysis and other useful intermediate files. These include the provided input files, the base pair (step) parameter files from which the models were generated, various bend and 3DNA analysis files and the 3D-DART log file.

In addition, the server provides a few convenience functions to further customize the output. These include options to change the PDB markup such as changing chain ID and renumbering residues. The server also offers the option to output the structures in a format and notation consistent with the macro-molecular docking program HADDOCK (14,15). In that case, an additional restraint file is generated to maintain the DNA conformation during the flexible refinement stage of the docking. These functionalities are actually used by the HADDOCK web server (http://haddock.chem.uu.nl/haddock) to automatically process DNA/RNA input structures.

REFERENCES

1. Flick,K.E., Jurica,M.S., Monnat,R.J. Jr and Stoddard,B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, **394**, 96–101.
2. Patikoglou,G. and Burley,S.K. (1997) Eukaryotic transcription factor-DNA complexes. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 289–325.
3. Allemann,R.K. and Egli,M. (1997) DNA recognition and bending. *Chem. Biol.*, **4**, 643–650.

4. Dixit,S.B. and Beveridge,D.L. (2006) Structural bioinformatics of DNA: a web-based tool for the analysis of molecular dynamics results and structure prediction. *Bioinformatics*, **22**, 1007–1009.
5. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
6. Macke,T. and Case,D.A. (1998) In Leontes,N.B. and SantaLucia,J.J. (eds), *Molecular Modeling of Nucleic Acids.*, American Chemcial Society, Washington DC, pp. 379–393.
7. Munteanu,M.G., Vlahovicek,K., Parthasaraty,S., Simon,I. and Pongor,S. (1998) Rod models of DNA: sequence-dependent aniso-tropic elastic modelling of local bending phenomena. *Trends Biochem. Sci.*, **23**, 341–346.
8. Tung,C.S. and Carter,E.S. II (1994) Nucleic acid modeling tool (NAMOT): an interactive graphic tool for modeling nucleic acid structures. *Comput. Appl. Biosci.*, **10**, 427–433.
9. Dickerson,R.E. (1998) DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.*, **26**, 1906–1926.
10. Dixit,S.B. and Beveridge,D.L. (2005) Axis curvature and ligand induced bending in the CAP-DNA oligomers. *Biophys. J.*, **88**, L04–L06.
11. El Hassan,M.A. and Calladine,C.R. (1998) Two distinct modes of protein-induced bending in DNA. *J. Mol. Biol.*, **282**, 331–343.
12. Jones,S., Heyningen,P., Berman,H. and Thornton,J.M. (1999) Protein-DNA interactions: a structural analysis. *J. Mol. Biol.*, **287**, 877–896.
13. Dickerson,R.E. (1989) Definitions and nomenclature of nucleic acid structure parameters. *J. Biomol. Struct. Dyn.*, **6**, 627–634.
14. Dominguez,C., Boelens,R. and Bonvin,A.M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **125**, 1731–1737.
15. de Vries,S.J., van Dijk,A.D., Krzeminski,M., van Dijk,M., Thureau,A., Hsu,V., Wassenaar,T. and Bonvin,A.M. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, **69**, 726–733.
16. van Dijk,M., van Dijk,A.D., Hsu,V., Boelens,R. and Bonvin,A.M. (2006) Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. *Nucleic Acids Res.*, **34**, 3317–3325.
17. Chandrasekaran,R.A. and Arnott,S. (1989) In Saenger,W. (ed.), *Landolt-Börnstein Numerical Data and functional Relationships in Science and Technology*. Vol. VII/1b, Springer-Verlag, pp. 31–170.