

# 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures

Xiang-Jun Lu<sup>1,2</sup> & Wilma K Olson<sup>1</sup>

<sup>1</sup>Department of Chemistry and Chemical Biology and BioMaPS Institute for Quantitative Biology, Rutgers—The State University of New Jersey, Piscataway, New Jersey 08854-8087, USA. <sup>2</sup>Department of Biological Sciences and Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10027, USA. Correspondence should be addressed to X.-J.L. (3dna.lu@gmail.com) or W.K.O. (wilma.olson@rutgers.edu).

Published online 3 July 2008; doi:10.1038/nprot.2008.104

**We present a set of protocols showing how to use the 3DNA suite of programs to analyze, rebuild and visualize three-dimensional nucleic-acid structures. The software determines a wide range of conformational parameters, including the identities and rigid-body parameters of interacting bases and base-pair steps, the nucleotides comprising helical fragments, the area of overlap of stacked bases and so on. The reconstruction of three-dimensional structure takes advantage of rigorously defined rigid-body parameters, producing rectangular block representations of the nucleic-acid bases and base pairs and all-atom models with approximate sugar-phosphate backbones. The visualization components create vector-based drawings and scenes that can be rendered as raster-graphics images, allowing for easy generation of publication-quality figures. The utility programs use geometric variables to control the view and scale of an object, for comparison of related structures. The commands run in seconds even for large structures. The software and related information are available at <http://3dna.rutgers.edu/>.**

## INTRODUCTION

In addition to the genetic message, DNA base sequence carries structural information important to its biological packaging and processing. This structural code governs how the double-helical molecule deforms in response to proteins and when and where the genetic information is expressed. For example, the binding of DNA onto the surface of the nucleosome (the protein–DNA assembly that constitutes the basic packaging unit in the nucleus) involves an indirect structural response dependent on the ease of deforming the constituent base-pair steps<sup>1</sup>. Most regulatory proteins and many enzymes, by contrast, recognize specific, sequence-dependent structural features of individual base pairs, making direct contact with DNA atoms unique to their binding sites<sup>2</sup>. Other enzymes that perform cutting or sealing operations of the phosphodiester backbone take advantage of the natural susceptibility of particular sequences to switch from the B to the A helical form<sup>3</sup>, binding to sites on DNA normally hidden within the B-type duplex. Thus DNA is not just a passive substrate of cellular proteins but an active player with physical properties capable of influencing the three-dimensional organization of genetic sequences and the activity of regulatory proteins and processing enzymes.

The diverse functions of RNA, ranging from the transmission of genetic information to chemical catalysis<sup>4</sup>, similarly reflect the sequences and structures of the heterocyclic base side groups. In contrast to double-stranded DNA, RNA molecules are generally single-stranded, with the ribonucleotide building blocks assembling into intricate, functionally specific, three-dimensional forms. The RNA chain folds back on itself, with double-helical elements interspersed between unpaired bulges, internal loops and multi-armed junctions, and the structure as a whole stabilized by tertiary interactions and/or bound ligands<sup>4</sup>. Whereas the complementary strands of double-helical DNA associate almost exclusively through canonical, Watson–Crick base pairing, RNA incorporates both Watson–Crick and noncanonical base

pairs, including interactions of the bases with the 2'-hydroxyl group unique to the ribose sugar<sup>5,6</sup>.

Growing interest in understanding the relationship between the global folding of nucleic acids and the sequence-dependent arrangements of the constituent bases and base pairs has stimulated the development of new approaches to analyze and depict DNA and RNA structures<sup>7,8</sup>. The spatial configurations of both associated bases and neighboring bases or base pairs depend upon the sequence<sup>9,10</sup>. Clear differences arise between the six rigid-body parameters (three angular variables and three translational variables) that describe the arrangements of different bases and base pairs (**Fig. 1**). The arrangements of paired bases are defined with respect to the orientation and displacement of an ideal, planar Watson–Crick base pair<sup>11</sup>, where the rigid-body parameters (buckle  $\kappa$ , propeller  $\pi$ , opening  $\sigma$  angles; shear  $S_x$ , stretch  $S_y$ , stagger  $S_z$  displacements<sup>7</sup>) are zero. The configurations of successive bases or base pairs are specified by an analogous set of variables: two bending angles (roll  $\rho$  and tilt  $\tau$ ), a rotation angle (twist  $\omega$ ), two in-plane dislocations (shift  $D_x$  and slide  $D_y$ ) and a vertical displacement (rise  $D_z$ )<sup>7</sup>. The local geometry of each dinucleotide step can also be described by another set of six helical parameters ( $x$ -displacement  $dx$ ,  $y$ -displacement  $dy$ , helical rise  $h$ , inclination  $\eta$ , tip  $\theta$  and helical twist  $\Omega_h$ ) that describe the regularity of the helix in which it occurs.

In contrast to the analysis of the Watson–Crick base pairs in double-helical DNA structures, the study of 'globular' RNA and other complex nucleic-acid structures requires preliminary knowledge of the bases comprising the helices, loops and folds of the polynucleotide chains. Although such information is self-evident in the spatial arrangements of small molecules, the sheer enormity of structures like the ribosome (2,876 nucleotides in the large 50S subunit<sup>12</sup>) hampers direct visual identification of secondary structural motifs, non-Watson–Crick pairs and long-range interactions.

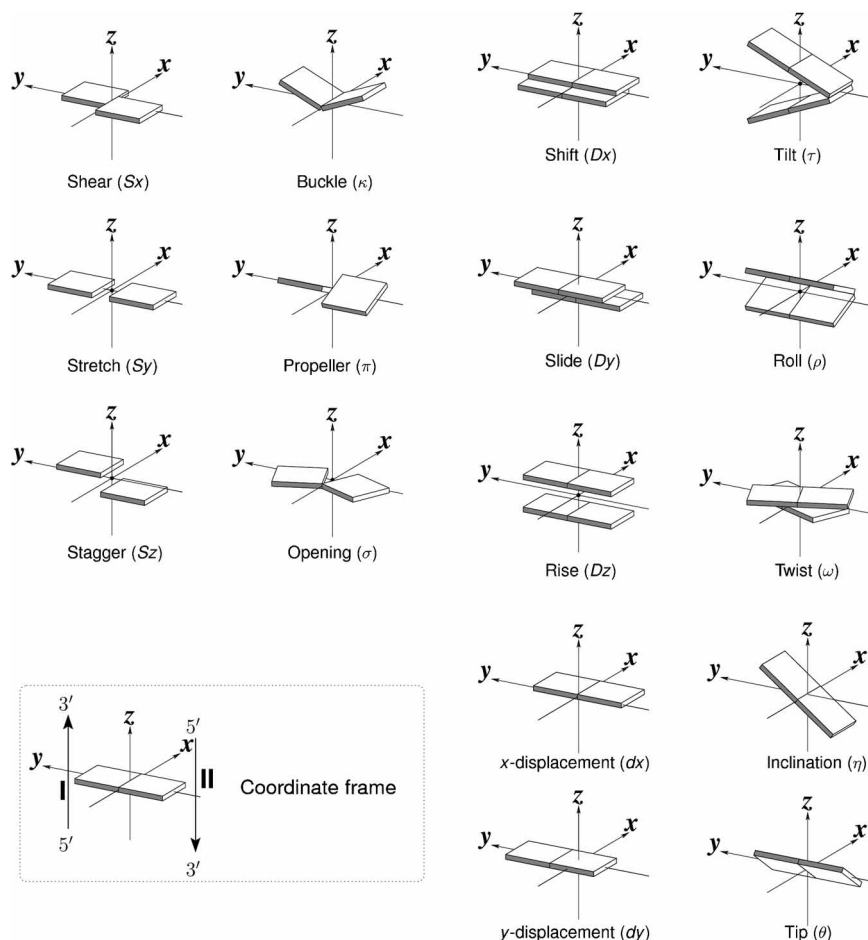
Studying such complicated nucleic-acid structures requires automated procedures to identify the short- and long-range base associations.

The arrangements of bases in nucleic-acid structures are closely tied to the immediate chemical environment in which they occur. That is, the deformations of helical structure occur in the context of the water molecules, proteins and other species in direct contact with the double helix. Previously, analyses of such interactions were literal and statistical in the sense of counting the number of different contacts in known structures<sup>13,14</sup>, or protein-centric in the sense of cataloging the various amino acids and folding motifs in close contact with nucleic acid<sup>15–18</sup>. Many of the tools developed to visualize protein–DNA complexes, for example, NUC-PLOT<sup>19</sup>, similarly focus on the identities of amino acids or secondary structural elements with respect to DNA primary structure. Understanding the effects of solvation on nucleic-acid sequence and structure requires a DNA- or RNA-centric perspective, which takes account of the precise locations of solvent, including closely associated protein and drug atoms, with respect to the heterocyclic bases or nucleotide steps<sup>20–23</sup>.

### Software overview

3DNA is a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid-containing structures, including their complexes with proteins and other ligands<sup>24</sup>. At its core, the software uses a simple matrix-based scheme<sup>25–29</sup> for calculating the complete set of rigid-body parameters that characterize the orientation and displacement of the base pairs, base-pair steps and single-stranded nucleotide steps that make up DNA and RNA structures. The description of structure is geometrically straightforward and the computation of parameters is mathematically rigorous<sup>24,28,29</sup>, allowing for *exact* rebuilding of a molecular structure based on the derived parameters. The software evolved from our efforts to resolve the discrepancies among seven commonly used nucleic-acid conformational analysis programs<sup>8,30</sup>, and makes use of the standard base-centered reference frame<sup>11</sup> recommended by the 1999 Tsukuba Workshop and later approved by the Nomenclature Committee of IUBMB (NC-IUBMB)/IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN).

3DNA is written in strict ANSI C computer language (~25,000 lines of code), with connecting Perl scripts. The software package consists of over 30 executable programs that can be run directly from the command line in a Unix/Linux environment (including Cygwin on Windows). Running each program with the ‘-h’ option (e.g., `find_pair -h`) provides the user with detailed usage



**Figure 1** | Pictorial definitions of rigid-body parameters used to describe the geometry of complementary (or noncomplementary) base pairs and sequential base-pair steps. The base pair reference frame (lower left) is constructed such that the *x* axis points away from the (shaded) minor-groove edge of a base or base pair and the *y* axis points toward the sequence strand (I). The relative position and orientation of successive base-pair planes are described with respect to both a dimer reference frame (upper right) and a local helical frame (lower right). Images illustrate positive values of the designated parameters. Reproduced from ref. 24 with permission from *Nucleic Acids Res.*

information and worked examples. Once the user is familiar with the package, it is straightforward to combine the various pieces and other command-line-driven (or capable) tools in a script to automate commonly repeated tasks, as demonstrated by the examples detailed herein.

The 3DNA software was first made available on the Internet in late 1999, and the currently used version (v1.5) was released in late 2002. An article published in 2003 (ref. 24) describes the major functionality available in the software. The 3DNA website includes unpublished technical details, a user’s manual and answers to frequently asked questions (FAQ). Over the years, we have answered hundreds of 3DNA-related questions from users all over the world, initially via email and more recently through the 3DNA forum, and have taken each and every query as an opportunity to improve the functionality of the software. We strive to respond to users as quickly and concretely as possible, often with a step-by-step recipe, until an issue is resolved. This process has helped us to refine and improve the software and has prompted us to add new functionality. As a result, 3DNA has found widespread use and the 3DNA homepage is linked from the websites of various



bioinformatics databases and computer-graphics tools, such as the Nucleic Acid Database (NDB)<sup>31</sup>, the RNA World website at Jena<sup>32</sup> and the Raster3D graphics toolset<sup>33</sup>.

The widespread use of 3DNA in the scientific community is evident from the rapidly growing number of citations of the 2003 article<sup>24</sup>. Although it is not practical to summarize the literature on all of the applications of the suite of programs, we draw attention here to several projects where 3DNA has played a particularly significant role: (i) the NDB<sup>31</sup> reports 3DNA base-pair and base-pair-step parameters for the double-helical fragments of each structure in the database; (ii) both the NDB<sup>31</sup> and the PDB (Protein Data Bank)<sup>34</sup> report the simple and informative molecular images generated with the `bloview` component of 3DNA; (iii) the nucleic-acid Solvation Web Service<sup>35</sup> uses 3DNA to find base pairs, collect conformational patterns and set the orientation of molecular images; (iv) the DNA structural bioinformatics server<sup>36</sup> (MDDNA) uses the 3DNA `rebuild` routine to construct models of arbitrary DNA sequences based on predictions obtained from molecular-dynamics simulations; (v) the ARTS<sup>37</sup> (Alignment of RNA Tertiary Structures) software uses 3DNA to identify both canonical and noncanonical base pairs for its seed-match construction; (vi) the information-driven protein–DNA docking method HADDOCK<sup>38</sup> uses 3DNA to build canonical B-DNA models, generate custom DNA libraries, calculate various base-pair and base-pair-step parameters, determine the torsion angles of the sugar–phosphate backbone and evaluate the puckering of the sugar ring; (vii) the base-pair server from this laboratory uses `find_pair` and other utility programs from 3DNA to build a comprehensive database and to generate illustrative molecular images of the base pairs in RNA structures; (viii) our recent identification of a roll-slide-induced DNA folding mechanism in chromatin<sup>1</sup> takes advantage of the rigorous matrix-based analysis and rebuilding functions of 3DNA and (ix) our generalization of the 3DNA code to treat the rigid-body parameters of planar molecular fragments<sup>39</sup> facilitates the classification of small, flexible drug molecules and identification of representative structures for three-dimensional quantitative structure-activity relationship analyses.

### Relationship to other programs

3DNA originated in the SCHNAaP/SCHNArP software<sup>28,29</sup> written to implement the computation of DNA rigid-body parameters with the mathematical formalism (CEHS) developed by El Hassan and Calladine<sup>27</sup> and to rebuild structures from those parameters, including the generation of ‘standardized’ base-stacking diagrams. Our efforts to resolve the discrepancies among nucleic-acid conformational analysis programs<sup>8,30</sup> and the definition of a standard base reference frame by the structural biology community<sup>11</sup> prompted us to take advantage of various features in the earlier programs and to adopt the standard frame in 3DNA. Specifically, we use the least-squared fitting procedure, developed by Babcock *et al.*<sup>40</sup> in RNA (Running Nucleic Acids), to superimpose a standard base<sup>41</sup> on an experimental structure in the determination of local base reference frames, and the procedure, employed in Newhelix<sup>42</sup>, to fit a global linear axis to a helix that does not deviate substantially from a regular structure.

The computation of DNA dimeric parameters is sensitive to the way in which a reference frame is assigned to each base and base pair<sup>8</sup>. As the reference frames used in 3DNA roughly coincide with

the frames used in CompDNA<sup>43</sup> and Curves<sup>44,45</sup>, the 3DNA base-pair and base-pair-step parameters closely resemble the values determined with CompDNA and the local parameters from Curves (except when DNA is highly kinked and the Curves values tend to be larger<sup>8</sup>). The discrepancies are more pronounced with programs, such as SCHNAaP/CEHS<sup>27–29</sup>, Newhelix/Freehelix<sup>42</sup>, and RNA<sup>40</sup>, which employ very different reference frames<sup>30</sup>. The `cehs` routine in 3DNA, nevertheless, allows the user to calculate authentic SCHNAaP/CEHS parameters, which (owing to the similar reference frames) closely match the values obtained with Newhelix/Freehelix, and with the ‘-r’ option within `cehs` to find authentic RNA<sup>40</sup> parameters. The 3DNA software also lets the user take advantage of the unique capability of Curves to find a global curvilinear helical axis and related global parameters, such as the widths and depths of the major and minor grooves, by providing an option ‘-c’ in the `find_pair` routine that generates an input file to Curves from a PDB file.

3DNA is unique among the many available nucleic-acid model-building programs in that it is purely geometrically based and that it allows for exact construction of a model from a set of straightforward and well-understood user-supplied parameters. A number of other programs make use of rigid-body methods, however for other purposes: NAB<sup>46</sup> (Nucleic Acid Building) provides ‘a programming environment for geometric and force-field manipulations of nucleic acids’; NAMOT<sup>47</sup> (Nucleic Acid Modeling Tool) builds and manipulates models based on a set of reduced coordinates developed by the authors; and Mc-Sym<sup>48</sup>, a knowledge-based software, builds RNA structures from molecular subfragments.

The visualization components in 3DNA yield drawings in vector-based Encapsulated PostScript (EPS) and Xfig formats. The latter can be manipulated easily using the free Xfig interactive drawing tool. The software also creates Raster3D scenes that can be fed directly into `render` or PyMOL for high-quality graphic images. RasMol<sup>49</sup> is especially effective for interactive visualization of the rectangular schematic presentations generated in Alchemy format with 3DNA. Thus, 3DNA complements many commonly used visualization tools (Raster3D, RasMol, MolScript and PyMOL).

3DNA is tailored for automatic visualization of structural variations in nucleic-acid base-pair geometry, a feature not found in other interactive and schematic drawing programs, such as DRAWNA<sup>50</sup> and UCSF Chimera<sup>51</sup>. 3DNA has built-in capabilities to generate and modify the sizes and shapes of Calladine–Drew-style base and base-pair ‘blocks’, and to highlight the minor-groove edges of such base or base-pair representations. As emphasized here and seen from the images of nucleic-acid containing structures at the NDB/PDB websites, the `bloview`-generated images are highly effective in revealing the key features of small- to medium-sized nucleic-acid structures and are used to represent each nucleic-acid containing structure in both the NDB<sup>31</sup> and PDB<sup>34</sup>. The programmatic generation of publication-quality diagrams of the base-stacking and higher-order base associations is also unique to 3DNA. Recently added tools for generating images of the ensembles of nucleic-acid structures found with nuclear magnetic resonance (NMR) methods and for converting ‘.r3d’ files for PyMOL ray-traced images are expected to gain wide use.

### Strengths and limitations of 3DNA

3DNA has many unique features that make it a practical tool in various real-world applications. For example, the automatic

**Figure 2** | Schematic diagrams of three representative rigid-body parameters. (a) Propeller ( $45^\circ$ ); (b) rise ( $3.4 \text{ \AA}$ ); (c) inclination ( $36^\circ$ ). Images in the left column were created automatically with 3DNA, and those on the right side are final, touched-up versions generated with Xfig. The six images were merged and aligned, and the composite figure shown here was exported into EPS format and subsequently converted to PNG format at 300 DPI using Ghostscript. Note that, with the automatic generation procedure in 3DNA, one can control the magnitude and sign of the parameters precisely and specify the orientation and scale of the image.

detection of all possible base pairs, higher-order base associations and helical fragments using `find_pair` makes analyzing nucleic-acid structures with 3DNA straightforward. To our knowledge, there are no other tools with such functionality. The real power of 3DNA, however, lies in its integrated approach of combining the analysis, reconstruction and visualization of three-dimensional nucleic-acid structures in a single software package. The examples given herein exemplify some of the possible applications. Except for some final touches or format conversion for use of commonly used, freely available software (RasMol, Raster3D, MolScript, Xfig, Ghostscript and ImageMagick), all the images in this article are generated directly from the program using scripts to automate an otherwise tedious, error-prone process.

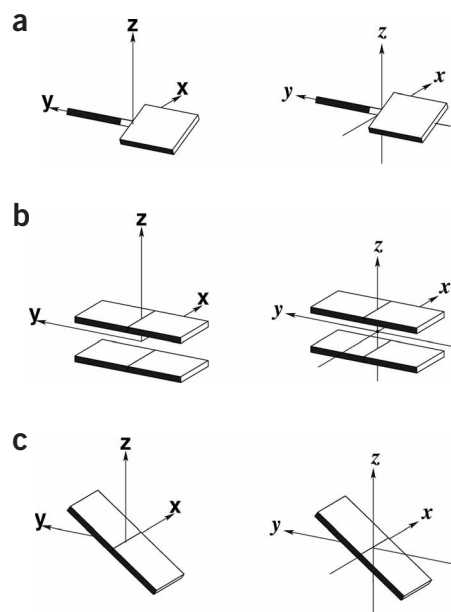
As with any software, 3DNA has become more sophisticated and robust over the years as it becomes more widely used, with user-reported bugs fixed, the code internally refined and new functionality added. The interactions with users from different backgrounds have given us the incentive to adapt the programs for further applications in related fields, for example, RNA structure-motif identification and alignment, structural analysis of DNA-protein complexes and modeling RNA folds. The reference-frame-based description of three-dimensional spatial geometry makes the methodology and algorithms in 3DNA directly applicable to these problems and treats them in a rigorous and consistent fashion. Similar extension and application of 3DNA to the growing number of DNA quadruplex structures should contribute to the analysis and understanding of these novel forms.

On the other hand, the command-line-driven style of 3DNA, although flexible and efficient, has a downside in that it is not very user-friendly, especially for educational purposes and to novices (non-Linux/Unix users). Our next goal is to create a Web-based interface for some commonly used features of the package, for example, generation of fiber models of arbitrary sequences, creation of `blockview` images, analysis of a structure, generation of stacking diagrams and so on. We will continue to support the 3DNA forum and maintain the online documentation to get the community involved in providing better documentation and more working examples.

### Sample applications of 3DNA

This procedure describes five protocols for the following: (A) preparation of schematic diagrams of representative base-pair parameters; (B) determination of DNA curvature associated with different roll distributions; (C) analysis and visualization of a DNA structure with a B-Z junction; (D) automatic identification of double-helical regions in a DNA-RNA junction; and (E) identification of higher-order base associations in ribosomal RNA.

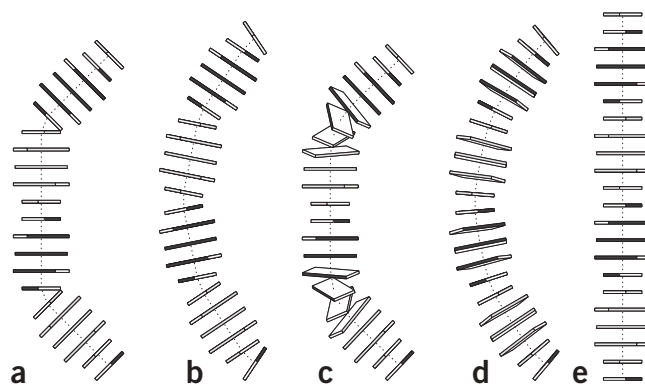
**(A) Preparation of schematic diagrams of representative base-pair parameters.** Although it is tedious, it is possible to obtain reasonable quality schematic images (typically rectangular blocks) of the rigid-body parameters describing the nucleic-acid base pairs



and base-pair steps with any modern computer-graphics tool; however, it is not trivial to control the relative position and orientation of the rectangular blocks precisely, to set the exact viewpoint and to maintain the same viewpoint and scale in different diagrams. Although the many representations in the literature of nucleic-acid rigid-body parameters, such as the propeller twist of complementary base pairs, the roll at deformed base-pair steps and so on, may look similar, they differ subtly from one another. 3DNA provides the necessary tools to generate such schematics more professionally by allowing for exact control of the conformational parameters as well as for the viewpoint and scale of the images (see, e.g., Figures 1 and 4 in ref. 24). Because many of the recent users of 3DNA have expressed an interest in illustrations such as those described in our earlier work<sup>24</sup> (reproduced here in Fig. 1), we first present the detailed steps needed to generate block images of each of the three types of rigid-body parameters, using propeller as a representative base-pair parameter, rise as a representative local base-pair-step parameter and inclination as a representative local helical parameter (Fig. 2).

**(B) Determination of DNA curvature associated with different roll distributions.** DNA is not a monotonous, regular straight helix, but rather deforms in response to external stresses in a sequence-dependent manner<sup>9,52–56</sup>. For example, the catabolic activator protein introduces two sharp kinks at pyrimidine-purine (YR) steps in its binding site, bending the DNA at these steps into the major groove via large positive roll<sup>57,58</sup>. By contrast, the DNA wrapped tightly around the assembly of histone proteins on the surface of the nucleosome core particle deforms via both positive and negative roll, with pronounced bending into the minor groove occurring at regularly spaced YR steps<sup>59</sup>. In between these two extremes of bending are other possible modes of deformation, such as the patterns of roll in the idealized curved DNA pathways presented in the Calladine *et al.*<sup>60</sup> textbook on DNA structural mechanics. The complicated three-dimensional nature of local bending makes it difficult, even for a seasoned scientist, to visualize how the variation in roll at different dinucleotide steps leads to global curvature. With 3DNA, one simply prepares a data file with

**Figure 3** | 3DNA-generated images of 22-bp DNA duplexes with the same overall 45° curvature per helical turn. (a–d) Four different theoretical distributions of roll angle achieving such curvature, according to models in the textbook by Calladine *et al.*<sup>60</sup>, namely two steps with 45° roll separated by a complete (10 bp) helical turn, five bends with 22.5° roll spaced at increments of a half helical turn and with alternating sign, two tracts of five 14°-rolled steps separated by a half helical turn and a smoothly curved structure with a repeating sinusoidal-like variation in roll (–9.00°, –7.28°, –2.78°, 2.78°, 7.28°, 9.00°, 7.28°, 2.78°, –2.78°, –7.28°); (e) ideal B-DNA reference structure with uniform (zero) roll.



any prescribed set of parameters and builds the structure to see what it looks like (Fig. 3). The matrix-based scheme adopted in the 3DNA analysis/rebuilding programs makes this a completely reversible and rigorous process. The second protocol shows how one can mimic the idealized global fold of a 22 base-pair fragment of DNA illustrated in Figure 4.8(a) in the textbook by Calladine *et al.*<sup>60</sup>.

**(C) Analysis and visualization of a DNA structure with a B–Z junction.** From a purely structural point of view, right-handed B-DNA has a positive twist (+36°, on average, per base-pair step) and left-handed Z-DNA has a negative twist angle (–60° per CGC/GCG repeating unit). Importantly, compared with B-DNA, the base pairs in Z-DNA are inverted along the long axis<sup>28</sup> (Fig. 4). Although various B–Z junction models have been offered for decades<sup>61,62</sup>, the first crystal structure of such a junction, in the presence of the human editing enzyme, double-stranded RNA-specific adenosine deaminase, was only recently solved by the Rich group at the Massachusetts Institute of Technology<sup>63</sup>.

The 3DNA analysis routine (`analyze`) checks for possible base-pair inversion within a double helix and classifies such structures as ‘unusual’ in the main output file. Before the publication of the B–Z junction article, we worked with the lead author on the analysis of a partial structure containing only the two base pairs at the B–Z junction, using the opportunity to update the 3DNA code to handle such structures automatically (Sung Chul Ha, personal email communication). For meaningful parameters to be calculated, the program reverses the *x* and *z* axes of the Z-DNA base pair at the junction (in the same way that the program must reverse the *y* and *z* axes of bases on complementary strands to compute sensible values of the base-pair parameters describing a Watson–Crick pair). Following publication of the B–Z junction structure, we were able to check the complete structure and verify that the program modification works properly. The third protocol shows how one can analyze and illustrate this interesting structural feature.

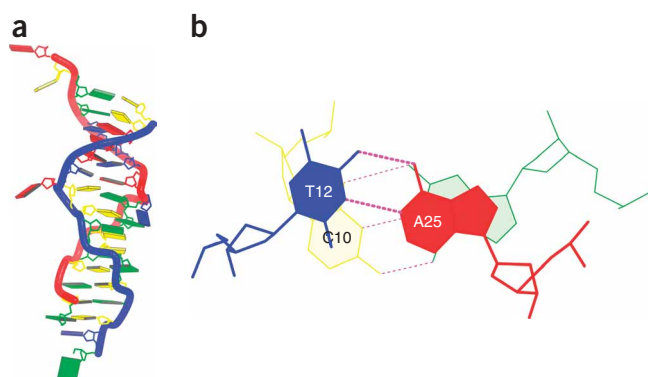
**(D) Automatic identification of double-helical regions in a DNA–RNA junction.** By taking advantage of the standard base reference

frame and selected geometric features, the `find_pair` program within 3DNA can identify all possible nucleic-acid base pairs, whether they are canonical Watson–Crick or noncanonical pairs and are made up of normal or modified bases, in any tautomeric form or protonation state. The program identifies helical regions from a purely base-stacked perspective by tracing each base pair against its neighborhood vertically, that is, in a direction perpendicular to the base-pair plane, regardless of whether there is a direct backbone connection between stacked base pairs. This program is the key application that makes the analysis of arbitrary nucleic-acid structures an easy job with 3DNA. Previously, one had to identify each base pair and each helical region manually or semiautomatically, a tedious and error-prone task, especially for RNA structures and DNA–protein or RNA–protein complexes.

The crystal structure of the RNA–DNA four-way junction (PDB entry 1EGK) solved by Nowakowski *et al.*<sup>64</sup> serves as an excellent example of the capabilities of 3DNA to identify base pairs and helical regions automatically and to draw a cylinder to visualize each relatively straight helix within the structure. The fourth protocol outlines the steps for finding the helical regions and generating a color-coded image of the overall structure (Fig. 5).

**(E) Identification of higher-order base associations in ribosomal RNA.** The same `find_pair` program used for locating base pairs and helical regions (option D) also identifies triplets, quadruplets and higher-order base associations. Starting from the identified base pairs, the program searches horizontally in the planes of the associated bases for further connections/interactions and checks for pairwise ‘compatibility’ to eliminate unreasonable associations. The program then collects the hydrogen bonds (H-bonds) formed from the nitrogen and oxygen atoms on the base and sugar–phosphate backbones of the contacted residues, including bifurcated H-bonds.

The last protocol demonstrates the capability of 3DNA to identify and generate publication-quality images of the two pentaplets, that is, associations of five paired bases, found in the fully refined crystal structure of the *H. marismortui* large ribosomal subunit<sup>65</sup> (PDB entry 1JJ2; Fig. 6).



**Figure 4** | Images of a B–Z-junction structure<sup>63</sup>. (a) `bloview`-generated image highlighting the ‘melted’ and flipped-out A11–T26 pair and the inverted base pairs at the C10T12/A25G27 junction step in the middle of the duplex; (b) standard stacking diagram of the junction obtained directly from 3DNA showing the intuitively negative twist angle, the appreciable shift and slide, and the same-strand base overlap of stacked base pairs. Base color-coding conforms to the NDB<sup>31</sup> convention: A (red); C (yellow); G (green); T (blue); U (aqua).

**MATERIALS  
EQUIPMENT**

A computer running a Linux/Unix based operating system (such as one of those listed below)

**EQUIPMENT SETUP**

3DNA is a software product that is distributed in binary form for common Linux/Unix operating systems, including various flavors of Linux distribution (e.g., Ubuntu, SuSE, RedHat and so on), SunOS, SGI, Mac OS X and Cygwin on Windows. The package distribution includes a subdirectory named NP\_Recipes/, which contains all the scripts, associated data files and graphics images used in this manuscript. The new 3DNA homepage, located at <http://3dna.rutgers.edu/>, has links to the 3DNA forum, the online documentation, answers to FAQs, technical details and other related information. See **Box 1** for detailed instructions on the installation of 3DNA.

**Data format** Input for the analysis of a nucleic-acid structure is a PDB-formatted file listing the Cartesian coordinates and chemical features (atom type, residue name and chain identifier) of all atoms in the structure of interest (see the RCSB PDB website for a detailed description of the PDB format). The PDB files can be those determined experimentally (e.g., X-ray or NMR from the PDB or NDB), or from modeling studies (e.g., molecular dynamics simulations).

For rectangular block schematic representations of bases and base pairs, 3DNA makes use of the Alchemy file format commonly used to depict the three-dimensional structures of small molecules. This format allows for explicit assignment of the types of atoms and chemical bond linkages (single, double or aromatic) that make up a given structure. We adopt the Alchemy format in 3DNA to take advantage of the popular molecular graphics program RasMol<sup>49</sup>, which reads and writes information in this format. While no detailed, authoritative description is found on the Internet, as for the PDB format, it is easy to understand Alchemy format from a few examples (see the examples/calladine\_drew/ directory or load a PDB structure into RasMol and then write back its coordinates in Alchemy format).

Input for the generation of a three-dimensional structure uses a simple 3DNA-specific text file format: the first line contains the number of base pairs or nucleotides in the desired (double- or single-stranded) structure; the second contains either a 0 or a 1 to denote whether the set of rigid-body variables will be in the form of base-pair/nucleotide step parameters (slide, roll



**Figure 5** | 3DNA-generated view of a four-way DNA-RNA junction<sup>64</sup> (PDB entry 1EGK), highlighting, with cylinders, the two straight helical regions detected by the program and orienting the structure into a view perpendicular to the long axes of the cylinders. Backbones are color-coded by chain identifiers, cylinders by helical identity, and nucleotides by chemical type. Magnesium ions and the waters to which they are coordinated in the grooves of the structure are represented by red dots. See legend to **Figure 4** for base color-coding.

and so on) or helical parameters (*x*-displacement, inclination and so on), respectively; the third line allows the user to insert a comment for the header of the parameter list. The lines that follow, equal to the number of base pairs or nucleotides specified in line 1, contain (i) the identities of the base pairs or nucleotides (e.g., G-C or G) and (ii) either the six rigid-body variables that relate each base pair or nucleotide to its predecessor or the 12 parameters that detail the positioning of the bases in the given pairs and the spatial relationship of each pair to its predecessor, that is, both the six base-pair parameters and the six dimeric step parameters. The lines are arranged so that the base sequence



**BOX 1 | 3DNA INSTALLATION INSTRUCTIONS**

Setting up 3DNA is a straightforward process, facilitated by a Perl script. Essentially, two things need to be done: (i) set up an environment variable X3DNA to point to the location where 3DNA is installed on your local system so that the various parameters files can be found; (ii) put the 3DNA binary directory into your command-line search path so that you can run the 3DNA commands conveniently. Here, we use the binary distribution compiled on Ubuntu Linux with the tarball file named 'Linux\_X3DNA\_v2.0.tar.gz' as an example to illustrate the process.

To install 3DNA, first download the tarball from <http://3dna.rutgers.edu/> to a directory of your choice on your local computer, presumably your home directory, and then follow the instructions below:

1. `tar zxvf Linux_X3DNA_v2.0.tar.gz`

The command creates a directory named X3DNA, which contains a complete 3DNA distribution. The directory has several subdirectories, including bin/ with the entire C binaries and Perl scripts, and NP\_Recipes/, which contains all the data files, scripts and the images used in this work.

2. `cd X3DNA/bin`

The command changes the current working directory to the 3DNA binary directory.

3. `x3dna_setup`

The command will produce an output message on the screen, like the following (on an Ubuntu Linux box with a default 'bash' shell):

To run X3DNA, you need to set up the following:

- the environment variable X3DNA
- add \$X3DNA/bin to your command-line search path

for your "bash" shell, please add the following into ~/.bashrc:

```
-----
export X3DNA=/home/xiangjun/X3DNA
export PATH=/home/xiangjun/X3DNA/bin:$PATH
-----
```

To complete setting up 3DNA, simply follow the output from x3dna\_setup as shown above, that is, set the environment variable X3DNA and add \$X3DNA/bin into your command-line search path by copying and pasting the two lines between the dashed lines into your current command window for the current session, or into your startup script (e.g., ~/.bashrc) for a permanent effect. For verification, type `find_pair -h`. Detailed command-line help related to the find\_pair routine will be displayed on the screen (the same '-h' option can be used to obtain help with all of the 3DNA programs and Perl scripts stored in the bin/ directory).

can be read down the first column and rigid-body parameters occur in the order of the axis of operation, with displacement variables listed before angular parameters (i.e., shear, stretch, stagger, buckle, propeller, opening in the case of base-pair parameters; shift, slide, rise, tilt, roll, twist in the case of base-pair-step parameters; and *x*-displacement, *y*-displacement, helical rise, inclination, tip, helical twist in the case of helical parameters). The step parameters of the first base pair or nucleotide are naturally zero (given the lack of a preceding residue). Although the omission of pair parameters has no effect on unpaired nucleotides, missing (i.e., null) values of the base-pair parameters automatically fix the associated bases in ideal, planar Watson–Crick arrangements. Data in this form provide a short-hand way to generate structures in which each base pair is represented as a single rectangular block (see Step 1B(i) of the PROCEDURE).

The accompanying **Supplementary Data** online contain a sample Alchemy file of a base-pair rectangular block ('Block\_BP.alc') and two sample input files

for rebuild: one for the generation of a base pair with a 45° propeller angle ('propeller.dat', Recipe no. 1 of **Box 2**), and another one for the construction of a 22 base-pair DNA with 45° curvature per helical turn ('roll\_a.dat', Recipe no. 2 of **Box 2**). Other worked examples can be found in the NP\_recipes/ directory that are distributed with 3DNA. Checking the examples is the most effective way to understand the manner in which nucleic-acid structure is specified and to use the software effectively.

**Third-party software** 3DNA is designed to be as self-contained as possible with minimal (clearly delineated) connections to third-party software. The analysis and rebuilding parts of 3DNA are totally independent of any third-party tools. The seven external programs listed in **Box 3** complement the visualization components of 3DNA, helping to create all the figures reported in this work. All of these programs are freely available and highly valuable to research problems of the sort that we describe.

## BOX 2 | FIVE STEP-BY-STEP RECIPES ILLUSTRATING VARIOUS CAPABILITIES OF 3DNA

Recipe no. 1: command-line script to create a schematic image for propeller of 45°

```
01 \cp $X3DNA/config/block/Block_M.alc Block_R.alc
02 \cp $X3DNA/config/block/Block_M.alc Block_Y.alc
03 rebuild -block2 propeller.dat propeller.alc
04 frame_mol -l -m -g -L=8 ref_frames.dat propeller.alc propeller_mg.alc
05 rotate_mol -r=side_view.dat propeller_mg.alc propeller_sv.alc
06 alc2img -f -g -s=166 propeller_sv.alc propeller_sv.fig
```

Recipe no. 2: command-line script to create a 22 base-pair long schematic duplex structure with a 45° curvature per helical turn

```
01 rebuild roll_a.dat roll_a_raw.alc
02 frame_mol -m -11,12 ref_frames.dat roll_a_raw.alc roll_a_mm.alc
03 rotate_mol -r=roty90.dat roll_a_mm.alc roll_a_ok.alc
04 alc2img -f -l -s=128 roll_a_ok.alc roll_a.fig
05 alc2img -l -s=7 roll_a_ok.alc roll_a.eps
```

Recipe no. 3: command-line script to analyze and visualize a B–Z junction structure

```
01 get_part 2acj.pdb 2acj_dna.pdb
02 blocview -t=300 -i=2acj_dna.png 2acj_dna.pdb
03 find_pair 2acj.pdb 2acj.inp
04 analyze 2acj.inp
05 ex_str -8 stacking.pdb bz_step.pdb
06 stack2img -c -d -o -l -t bz_step.pdb bz_step.eps
```

Recipe no. 4: command-line script to illustrate the three helices in a four-way DNA–RNA junction

```
01 find_pair 1egk.pdb stdout
02 ex_str -1 hel_regions.pdb h1.pdb
03 ex_str -2 hel_regions.pdb h2.pdb
04 ex_str -3 hel_regions.pdb h3.pdb
05 find_pair h1.pdb stdout | analyze
06 find_pair h2.pdb stdout | analyze
07 find_pair h3.pdb stdout | analyze
08 rotate_mol -t=h1x_h3y.rot 1egk.pdb 1egk_h1x_h3y.pdb
09 find_pair 1egk_h1x_h3y.pdb stdout
10 ex_str -1 hel_regions.pdb h1x.pdb
11 ex_str -3 hel_regions.pdb h3y.pdb
12 find_pair h1x.pdb stdout | analyze
13 \mv poc_haxis.r3d h1x.r3d
14 find_pair h3y.pdb stdout | analyze
15 \mv poc_haxis.r3d h3y.r3d
16 blocview -o 1egk_h1x_h3y.pdb
17 cat t.r3d transparent.r3d h1x_ok.r3d h3y_ok.r3d > 1egk_ok.r3d
18 x3dna_r3d2png 1egk_ok.r3d 1egk_ok.png
```

Recipe no. 5: command-line script to identify pentaplets in the large ribosomal subunit

```
01 find_pair -p 1jj2.pdb 1jj2.bps
02 ex_str -117 multiplets.pdb p1.pdb
03 ex_str -135 multiplets.pdb p2.pdb
04 stack2img -s=300 -f -cdol p1.pdb p1.fig
05 stack2img -s=300 -f -cdol p2.pdb p2.fig
```

## BOX 3 | EXTERNAL PROGRAMS THAT COMPLEMENT THE VISUALIZATION COMPONENTS OF 3DNA

1. RasMol (<http://www.umass.edu/microbio/rasmol/>), although not directly used in generating the images made for this article, is a popular, easy-to-use program for molecular structure visualization and a useful product for gaining experience with Alchemy files (with options `'-alchemy'` and `'-noconnect'`).
2. MolScript (<http://www.avatar.se/molscript/>) is a program for displaying molecular three-dimensional structures in both schematic and detailed representations. It is used in 3DNA for generating a Raster3D input file for rendering protein and nucleic-acid backbone ribbons.
3. Raster3D (<http://skuld.bmsc.washington.edu/raster3d/>) is a set of tools for generating high-quality raster images. 3DNA makes use of `render` to generate raster-graphics images in Raster3D `'r3d'` format.
4. PyMOL (<http://pymol.sourceforge.net/>) is a user-sponsored molecular visualization system with scripting capabilities (`'pml'` script) used by 3DNA for integration with other command-line-driven tools. A user can also load the `'r3d'` files created by 3DNA into PyMOL to manipulate them interactively, in the same way that one would use the normal Raster3D/MolScript to PyMOL interface.
5. Ghostscript (<http://www.sourceforge.net/projects/ghostscript/>) is an interpreter for the Adobe PostScript language and can serve as a file-format converter. It is used by 3DNA for converting an EPS file to a PNG image at a specific resolution, 300 DPI for the figures in this article. Commercial products like Adobe Illustrator can substitute for Ghostscript for interactive uses.
6. ImageMagick (<http://www.imagemagick.org/script/index.php>) is a useful, general-purpose, image-processing toolset for creating, editing and composing bitmap images. 3DNA uses the `display` and `convert` components of the software. Adobe Photoshop can also substitute for ImageMagick for interactive uses.
7. Xfig (<http://www.xfig.org/>) is an interactive drawing tool chosen by 3DNA for its simplicity, quality and availability. The software fulfills our needs for line drawing and composing composite figures and also allows for direct editing of the Xfig files generated by 3DNA.

### PROCEDURE

**1|** In this protocol, we show five real-world applications of 3DNA, inspired largely by our interactions with users over the years. Combined together, they illustrate the versatile capabilities of 3DNA, made possible by its unique integrated approaches on the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures.

To prepare schematic diagrams of representative base-pair parameters, in particular, to generate an image of a base pair with a propeller of  $45^\circ$ , follow the steps in option A. For determining the curvature of DNA associated with different roll distributions, follow the steps in option B. To analyze and visualize a DNA structure with a B-Z junction, follow the steps in option C. For automatic identification of double-helical regions in a DNA-RNA junction, go to option D. To identify higher-order base associations in ribosomal RNA, go to option E.

#### (A) Preparation of schematic diagrams of representative base-pair parameters, using propeller of $45^\circ$ as an example (Recipe no. 1, Box 2)

- (i) Prepare a data file entitled `'propeller.dat'` with all base-pair parameters other than propeller ( $45^\circ$ ) set to zero in the simple 3DNA-specific text format described above (see data format in the MATERIALS). This file can be created and edited with any text editor (e.g., `vi` and `emacs`), or more easily done using the 3DNA utility program `regular_dna` with the number of repeating units set to unity (only one base pair). It should be noted that sequence identity does not matter in the construction of a general base-pair schematic.
- (ii) Assign blocks of the same size ( $4.5 \times 5.0 \times 0.5 \text{ \AA}$ ) to the complementary purine (R) and pyrimidine (Y) residues of a Watson-Crick base pair using a `'Block_M.alc'` file for both bases, that is, `'Block_R.alc'` and `'Block_Y.alc'` (lines 01 and 02). The edges of the blocks run parallel to the axes of the bases, that is,  $4.5 \text{ \AA}$  across the short axis (joining the minor- and major-groove edges) of each base,  $5.0 \text{ \AA}$  along the long axis (pointing toward the sugar-phosphate backbone) and  $0.5 \text{ \AA}$  along the normal. Note that the `'\'` character in front of `'cp'`, that is, `'\cp'`, forces the copy command to run noninteractively.
- (iii) Build the schematic structure of the base pair in Alchemy format, by running `rebuild` with the two-block option `'-block2'`, that is, using one rectangular block for each base and naming the resultant set of coordinates `'propeller.alc'` (line 03). The Alchemy file can be conveniently viewed with RasMol<sup>49</sup> (using the `'-alchemy'` and `'-noconnect'` options).
 

**! CAUTION** The option `'-noconnect'` makes sure that RasMol uses only the linkage information specified in the Alchemy file (by setting the `CalcBondsFlag` to false). It is not required for the `'propeller.alc'` file, but it is critical for the `'propeller_mg.alc'` (Step 1A(iv)) and `'propeller_sv.alc'` files (Step 1A(vi)). The latter two Alchemy files contain explicitly specified coordinate axes, which would interfere with the default bond-calculation algorithm in RasMol.
- (iv) Reorient the structure with respect to the base-pair reference frame (`'-1'`) using `frame_mol` and naming the transformed coordinate file `'propeller_mg.alc'` (line 04). The option `'-m'` specifies that the minor groove faces the viewer, and the options `'-g -L=8'` that the global reference axes are  $8.0 \text{ \AA}$  in length. Note that the viewing options, that is, `'-1 -m -g -L=8'`, can be listed in any order in 3DNA, in either lower or upper case (although `'-L'` for length is used here to avoid any confusion between a lower case `'l'` and the number one `'1'` introduced to specify the reference frame), and that single-letter



options can be bundled (e.g., '-m' and '-g' can be combined as either '-mg' or '-gm'). Strictly speaking, the '-1' option is redundant, since `rebuild` places the global reference frame by default on the first base-pair frame of a generated structure.

- (v) Create the data file 'side\_view.dat' containing the following two lines to transform the image so that the imposed propeller angle can be seen more clearly. The first line rotates the structure about the (vertical) *y* axis by  $-30^\circ$  and the second about the (horizontal) *x* axis by  $+20^\circ$ .

```
by rotation y -30
by rotation x 20
```

- (vi) Rotate (with `rotate_mol`) the base-pair block ('propeller\_mg.alc') using the preceding rotation file ('-r=side\_view.dat') to create the new output in side view ('propeller\_sv.alc', line 05).
- (vii) Create the final schematic image ('propeller\_sv.fig') in Xfig ('-f') format by applying `alc2img` to the above set of rotated coordinates ('propeller\_sv.alc'), with the global axes represented by arrows ('-g') and the scale factor set to a specified number ('-s=166') (line 06). The latter value is expressed in Xfig-defined units, where a value of 8,333 units corresponds, at a resolution of 1,200 units per inch, to a maximum figure dimension of 6.94 inches or 17.6 cm. The 3DNA image-generating programs (`alc2img`, `stack2img`, `r3d_atom`, `bloview` and so on) automatically output the currently used scale factor in the diagnostic messages for drawing images of different styles (PostScript, Xfig, Raster3D). Setting the scale factor to a fixed value with the '-s' option ensures that a series of related structures are drawn, as illustrated in **Figure 2** (propeller, rise and inclination), in exactly the same scale.

### (B) Determination of DNA curvature associated with different roll distributions, using two $45^\circ$ kinked steps separated by 10 base pairs as an example (Recipe no. 2, Box 2)

- (i) Prepare a data file entitled 'roll\_a.dat' specifying the base-pair-step parameters that relate each base pair to its predecessor in a 22 base-pair fragment. The base pairs in the structure span 2.1 double-helical turns of DNA, that is, 21 base-pair steps, each twisted by  $36^\circ$ . The overall curvature of  $45^\circ$  per helical turn is produced by two properly phased kinks at base-pair steps 6 and 16, each with a  $45^\circ$  roll angle. Thus, in the data file, the rise is set to 3.34 Å and the twist to  $36^\circ$  throughout the structure, and all remaining step parameters, other than the  $45^\circ$  roll at base-pair steps 6 and 16, are fixed at zero. The omission of base-pair parameters (e.g., buckle, propeller) in each line ensures that these values are all zero (see discussion of file format in MATERIALS). Here, each base pair is represented by a single ( $4.5 \times 10.0 \times 0.5$  Å) rectangular block.
- (ii) Build the structure (using `rebuild`) from the base-pair-step parameters specified in the above input file 'roll\_a.dat' and store the coordinates of the base-pair blocks in the output file named 'roll\_a\_raw.alc' (line 01). By default, `rebuild` constructs a structure using base-pair blocks ('-block1') in Alchemy format, as we want here.
- (iii) Place the molecular coordinate system in the middle frame of the central 11–12 dinucleotide step ('-11,12') by employing `frame_mol`, set the view so that the minor groove faces the viewer ('-m'), and name the resultant file 'roll\_a\_mm.alc' (line 02). It should be noted that the 'ref\_frames.dat' file, which contains the reference-frame information for each base pair, can be generated by a number of programs in 3DNA, including `find_pair`, `analyze` and `rebuild`. Here, this information comes from the previously executed `rebuild` step.
- (iv) Rotate the DNA (using `rotate_mol`) by  $90^\circ$  around the (vertical) *y* axis of the global coordinate frame ('-r=roty90.dat', where 'roty90.dat' is a file with the instruction 'by rotation y 90') so that the structure adopts the desired view ('roll\_a\_ok.alc') (line 03). As pointed out in option A above, the various Alchemy files generated with 3DNA can be conventionally viewed and manipulated (e.g., rotated) with RasMol<sup>49</sup>.
- (v) Create the final illustrated image 'roll\_a.fig' in Xfig format ('-f') using `alc2img`, with consecutive base-pair centers connected by dotted lines ('-l') (line 04). The fixed scale factor ('-s=128') makes sure that all five images (**Fig. 3**) are in exactly the same scale. Finally, for illustration purposes, the last line (05) generates the same image in EPS, the default image format from `alc2img`. The '-l' and '-s=7' options in this instruction have the same meaning as the 'l' and 's' symbols used above to generate an image in Xfig format.

### (C) Analysis and visualization of a DNA structure with a B–Z junction (Recipe no. 3, Box 2)

- (i) Extract (with `get_part`) the nucleic-acid part ('2acj\_dna.pdb') of the protein–DNA complex (PDB entry 2ACJ) (line 01).
- (ii) Use `bloview` to obtain a simplified, yet informative color-coded image ('2acj\_dna.png') of the DNA (**Fig. 4a**) (line 02). The recipe includes instructions for the creation of a PyMOL ray-traced image ('-i=2acj\_dna.png') at a resolution of 300 dots per inch (DPI) ('-t=300').
- (iii) Determine the base-pair identities and content directly from the original protein-bound DNA file ('2acj.pdb') using `find_pair` (line 03), generating the file ('2acj.inp') to be used as input for the analysis routine. Note that `find_pair` also outputs a diagnostic message '^v opposite bp direction: 1(14) 8(8)-9(9)', indicating that the *z* axes of base pairs 8 and 9 in this 14 base-pair double helix (helix 1, the only helical fragment in the structure) run in opposite directions, as characteristic of a B–Z junction.
- (iv) Calculate structural parameters using `analyze` from the base-pairing information in file '2acj.inp' (line 04), storing the output in the file '2acj.out'. The 'Structure classification' component of `analyze` correctly identifies the structure as

containing a junction located between a right-handed helix and a left-handed one. The analysis program also creates a fixed-named file called 'stacking.pdb' that stores the coordinates of all dinucleotide steps in terms of their middle frame.

- (v) Extract (using `ex_str`) the B-Z junction at base-pair step 8 ('-8') from the file 'stacking.pdb' and save the coordinates in the PDB-formatted file 'bz\_step.pdb' (line 05).
- (vi) Generate a standardized stacking diagram of the B-Z junction with `stack2img` (line 06). The completed image (Fig. 4b) is color-coded by residue ('-c'), with hydrogen bonds drawn ('-d'), base rings filled ('-o'), bases labeled ('-l') at the centers of the six-membered rings, and the base pair oriented in a top-down view ('-t') such that the (middle-frame) long axis of the step is horizontal and the leading strand lies on the left. As noted above, these five single-letter options can be combined as '-cdolt'.

**(D) Automatic identification of double-helical regions in a DNA-RNA junction (Recipe no. 4, Box 2)**

- (i) Using RasMol or any other molecular-graphics program (e.g., PyMOL), examine the structure of the DNA-RNA junction (PDB entry 1EGK) and confirm that it forms three helices.
- (ii) Locate the three helical regions (here termed h1, h2 and h3) in the 1EGK coordinate file ('1egk.pdb') using `find_pair` and store the helical content information in the fixed-named file 'hel\_regions.pdb' (line 01). Note that 'stdout' in the script stands for the standard output stream, normally connected to the screen, and that a fixed-named file 'hel\_regions.pdb' is generated by default and thus not given in the command line.
- (iii) Extract the atomic coordinates of the three helices into separate files ('h1.pdb', 'h2.pdb', 'h3.pdb') using `ex_str` (lines 02-04) and calculate the structural parameters of each with `analyze` (lines 05-07). The '|' is the pipe used to connect the standard output from `find_pair` ('stdout') directly into `analyze` without resorting to an intermediate data file (as in line 03 of Recipe no. 3). The 3DNA default settings, defined in the file 'misc\_3dna.par' (see Step 1E(i)), show that helices h1 and h3 are relatively straight, and thus a linear least-squared global helical axis is fitted to each. The analysis program also provides a normalized helical-axis vector for h1 and h3, and the two end points that each axis passes through, expressed in the original coordinate frame of the PDB file. Such information is useful for calculating the global bending angle between the two helices (86° here) and the shortest distance between the centers of the helices (20.2 Å), the latter value indicating that the two helices are nearly touching (given the characteristic 20 Å diameter of double-helical DNA/RNA).
- (iv) Reorient the structure as a whole ('1egk.pdb') with `rotate_mol` such that the two helical-axis vectors (associated with h1 and h3) lie in the x-y plane (line 08). Here h1 is set horizontally (along the x axis) and h3 is set vertically, with its component orthogonal to the (cylindrical) axis of h1 along the y axis. The z axis is chosen to complete a right-handed coordinate system. The above transformation is expressed in a transformation matrix file 'h1x\_h3y.rot', which is used with the '-t' option of `rotate_mol` ('-t=h1x\_h3y.rot') to transform the PDB data file '1egk.pdb' into the desired orientation in the file '1egk\_h1x\_h3y.pdb'.

**! CAUTION** Two warning messages, 'Warning: x- and y-axes are not orthogonal' and 'Warning: z-axis is not perpendicular to xy-plane', are not a matter of concern. The program (`rotate_mol`) checks if the three input axes are orthogonal and performs vector normalization/decomposition as necessary. Here, the h1 and h3 axes are not orthogonal, and the z-axis is randomly chosen and not necessarily perpendicular to the plane defined by the h1 and h3 axes. The content of the 'h1x\_h3y.rot' file is as follows:

```
1 # x-, y-, z-axes row-rise
0.000 0.000 0.000 # translation
-0.348 0.922 -0.167 # x: h1 axis, from h1.out
0.597 0.161 -0.785 # y: h3 axis, from h3.out
0.000 0.000 1.000 # z: can be anything
```

- (v) Repeat the process of identifying helical regions in the newly reoriented structure, that is, apply `find_pair` to the transformed PDB file '1egk\_h1x\_h3y.pdb', and extract (using `ex_str`) the two straight helices h1 and h3 into the files 'h1x.pdb' and 'h3y.pdb', respectively (lines 09-11).
- (vi) Recompute the structural parameters of the residues in the reoriented helix h1, that is, apply `find_pair` followed by `analyze` to 'h1x.pdb' (line 12). The output includes an auxiliary fixed-named file, 'poc\_haxis.r3d', which contains the Raster3D scenes for the cylinders with radii based on the mean displacement of the P, O4' and C1' atoms.
- (vii) Rename ('\mv') the Raster3D file 'poc\_haxis.r3d' into file 'h1x.r3d' (line 13) and manually modify 'h1x.r3d', by uncommenting (i.e., selecting) the cylinder scenes based on the phosphorus atoms, changing the color to red (by modifying its RGB color components). The updated Raster3D file is now called 'h1x\_ok.r3d'. Compare the two files ('h1x.r3d' and 'h1x\_ok.r3d') to see the difference, for example, use the Unix command `diff`.
- (viii) Repeat Steps 1D(vi) and 1D(vii) for the reoriented helix h3 to obtain the file 'h3y\_ok.r3d', with the color of the cylinder set to green (lines 14-15).



- (ix) Run `blocview` (line 16) on the previously transformed PDB file '1egk\_h1x\_h3y.pdb' to get the color-coded-block Raster3D scenes in the fixed-named file 't.r3d' using the original ('-o') coordinates (i.e., without automatically setting the 'best-view' based on the longest principal axes of the structure).
- (x) Employ `cat` (a simple Unix utility program) to combine all the 'r3d' files ('t.r3d' from Step 1D(ix), 'h1x\_ok.r3d' and 'h3y\_ok.r3d' from Steps 1D(vi–viii)) into a single file (by file redirection '>') called '1egk\_ok.r3d', including the 'transparent.r3d' scenes from Raster3D distribution that make the two cylinders transparent. The file 't.r3d' must be in the first position in the list for its header information, and 'transparent.r3d' should be second followed by 'h1x\_ok.r3d' and 'h3y\_ok.r3d' (in either order) to ensure transparency in the two cylinders (line 17).
- (xi) Transform the scenes into a PNG image ('1egk\_ok.png') with the utility Perl script `x3dna_r3d2png` (line 18) that employs the `render` program of Raster3D (**Fig. 5**).

**(E) Identification of higher-order base associations in ribosomal RNA (Recipe no. 5, Box 2)**

- (i) Find all pairwise interactions and higher-order associations ('1jj2.bps') in the *H. marismortui* large ribosomal subunit<sup>65</sup> ('1jj2.pdb') using `find_pair` with the '-p' option (line 01). The program generates a PDB coordinate file named 'multiplets.pdb', which contains all triplets and higher-order base associations (187 cases), each defined with respect to its middle reference frame. Specifically, the z axis of the assembly lies along the average of all the base normals, giving an objective, top-down view of the multibase association. Here, a primary H-bond distance cut-off of 3.3 Å, a bifurcated H-bond distance cut-off of 3.1 Å, and a vertical base-separation (stagger) cut-off of 2.0 Å are used by modifying the parameters file 'misc\_3dna.par'.
- (ii) Extract (with `ex_str`) the two pentaplets, numbered 117 and 135 in the file 'multiplets.pdb', into two separate PDB files ('p1.pdb' and 'p2.pdb', lines 02–03).
 

**! CAUTION:** The serial numbers given here for the two pentaplets, 117 and 135, depend on the parameter settings in 'misc\_3dna.par'. They will differ if, for example, the H-bonding distance cut-off is changed.
- (iii) Generate images of the pentaplets in Xfig ('-f') with options similar to those used in option C (Recipe no. 3 of **Box 2**) and a fixed scale factor ('-s=300') that ensures that the two molecular representations are scaled exactly the same way (lines 04–05).

**● TIMING**

With the software tools properly installed, rerunning each of the scripts mentioned above (on an ordinary Ubuntu Linux box) takes less than a minute, even for the large 50S ribosomal subunit. Most tasks take only a few seconds or less.

**ANTICIPATED RESULTS**

**(A) Preparing schematic diagrams of representative base-pair parameters**

The detailed procedure given in Step 1A (see Recipe no. 1 of **Box 2**) can be put into a shell script and run automatically to yield the desired image (**Fig. 2a**, left). With Xfig, it is then straightforward to touch-up the 3DNA-generated diagram ('propeller\_sv.fig') directly, e.g., changing the font, extending the axes and so on (**Fig. 2a**, right). The advantage of the procedure is the precise control of the value, viewpoint and scale of the illustrated parameter. The script makes it easy to set propeller to another value, for example, the  $-15^\circ$  value typical of the base pairs found in A- and B-DNA crystal structures<sup>11</sup>, and repeat the procedure to generate another block diagram.

The procedure needed to make a schematic diagram of rise (**Fig. 2b**) or inclination (**Fig. 2c**) is similar to that described for propeller, except for the replacement of the complementary bases by two successive base pairs. The two base blocks in each pair will touch each other if they are of the same size and if all base-pair parameters are zero, generating an image of size ( $4.5 \times 10 \times 0.5$  Å) with a line running along the short x axis through the middle of the base-pair block. The central line is actually the superimposed edges of the two base blocks that make up each base pair (**Fig. 2b,c**). The images of nucleic-acid parameters in Figure 1 of the 2003 paper<sup>24</sup>, reproduced here, also as **Figure 1**, were generated this way with an early version of 3DNA, where each of the illustrated displacement parameters in the figure, except for rise (3.4 Å), is 2 Å, and each of the angular parameters, except for inclination and tip (each  $36^\circ$ ), is  $45^\circ$ .

**(B) DNA curvature associated with different roll distributions**

Application of Step 1B yields an image that mimics the global DNA fold illustrated in Figure 4.8(a) in the textbook by Calladine *et al.*<sup>60</sup> (**Fig. 3a**). The overall curvature of  $45^\circ$  per helical turn—produced by two properly phased kinks at base-pair steps 6 and 16, each with a  $45^\circ$  roll angle—is striking as compared with a regular (unkinked) duplex of the same length (**Fig. 3e**).

This built-as-desired strategy for nucleic-acid structure generation, made possible with 3DNA, has broader implications in real-world applications. Firstly, one can easily modify the parameter file and rebuild a structure to see how selected local parameters affect global structure. For example, the images (**Figs. 3b–d**), created with the same procedure as shown in Step 1B (see Recipe no. 2 of **Box 2**) but incorporating different parameter files, show how different patterns of roll yield the same  $45^\circ$  global bend of DNA. On the other hand, although conventional models attribute the global fold of DNA to the bending

## PROTOCOL

and twisting of the constituent base-pair steps, a recent paper from our group<sup>1</sup> shows that simply by setting the slide, the rigid-body parameter that specifies the displacement of successive base pairs along their long axis (**Fig. 1**), to zero (typical of B-DNA<sup>11</sup>), the super-helical fold of nucleosomal DNA collapses to a circle. Indeed, we uncovered this novel roll-slide mechanism of DNA folding in chromatin by using the 3DNA *rebuild* routine to model nucleosomes. Secondly, *rebuild* allows for the construction of atomic-level nucleic-acid structures ('atomic' option) with sugar-phosphate backbones in pre-assigned, fixed conformations (specified by standard residue files; see the FAQ section at the 3DNA website).

Such models, which have precise base-pair geometry but approximate (sometimes distorted) backbone connections, provide a useful starting point<sup>66</sup> and basis for analysis<sup>36</sup> of all-atom simulations. Thirdly, as demonstrated in SCHNArP<sup>29</sup>, various DNA bending 'rules', with different sets of tilt/roll/twist values at the constituent dinucleotide or trinucleotide steps, can be easily incorporated within a script that transforms a sequence with an assigned 'bending model' into an input file that feeds into *rebuild*.

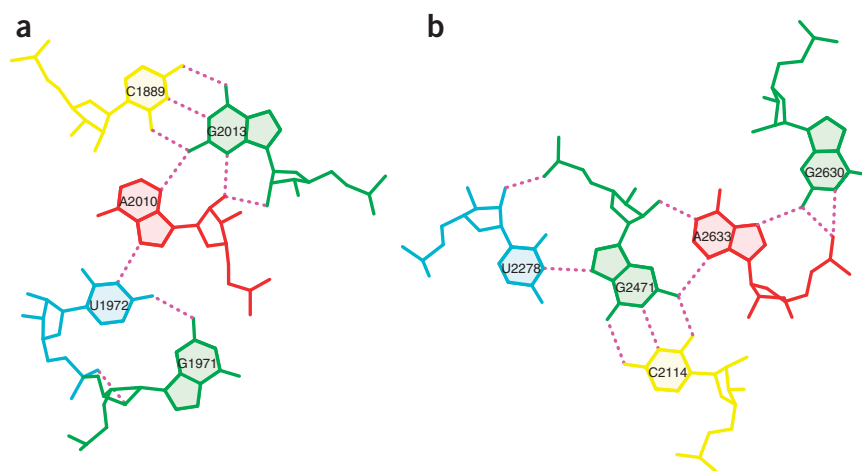
### (C) Analysis and visualization of a DNA structure with a B-Z junction

In the molecular image (**Fig. 4a**) created with Step 1C, the unpaired bases at the chain ends and the broken and flipped-out A-T base pair in the middle of the B-Z structure immediately stand out. Here A, C, G and T are colored by default in red, yellow, green and blue, respectively, following the NDB base color-coding convention. The stacked C10T12/A25G27 B-Z junction abuts the flipped-out A-T pair in the center of the duplex. Whereas the (long, unshaded) major-groove edge of the C10-G27 base pair in the lower Z-DNA half of the structure faces the viewer, the (long, shaded) minor-groove edge of the T12-A25 base pair in the upper B-DNA half faces outward. The obvious buckling of the two base pairs in the opposite sense leads to the small rise (2.6 Å) at the junction step<sup>11</sup>.

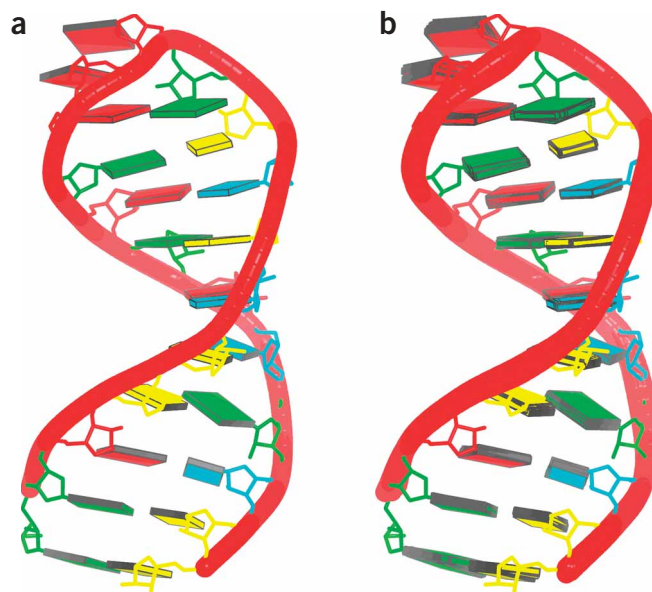
### BOX 4 | AUTOMATIC GENERATION OF SCHEMATIC IMAGES OF AN NMR ENSEMBLE OF AN RNA STEM-HAIRPIN

The newly added Perl script *nmr\_ensemble* generates an NMR-ensemble image (**Fig. 7b**) that combines the unique features of *blockview*, that is, the rectangular block representation of bases, color-coding of chains, residues and shading of base minor-groove edges, with each member of the structural ensemble automatically set in the principal-axis frame of the best representative structure (**Fig. 7a**). Here, we illustrate the set of solution structures, reported by Zoll *et al.*<sup>70</sup>, of variants of the Y domain of the poliovirus single-stranded RNA genome. We orient the best representative conformer in its most extended view (**Fig. 7a**) and record the conversion matrix for subsequent transformation of the remaining NMR conformers to the same frame.

The completed set of ensemble images (**Fig. 7b**) clearly shows the identities of the base residues within the double-helical stem and the unpaired adenines and noncanonical G-A pair in the single-stranded RNA hairpin at the top of the image. The structural variations of nucleotide residues (e.g., the larger, more floppy variation of the unpaired adenines versus the movement of the G-A pair) are also evident.



**Figure 6** | Two pentaplets from the refined crystal structure of the *H. marismortui* large ribosomal subunit<sup>65</sup> (PDB entry 1JJ2) automatically detected and illustrated with 3DNA. (a) C1889-G2013-A2010-U1972-G1971; (b) U2278-G2471-C2114-A2633-G2630. Note the direct involvement of the O2' sugar and O2P backbone atoms in H-bonding and the bridging role of the adenine. See legend to **Figure 4** for base color-coding.



**Figure 7** | Schematic *nmr\_ensemble*-generated images of the bases and backbones of the 3'-UTR Y-stem from the poliovirus RNA genome<sup>70</sup>. (a) The best-representative NMR structure; and (b) the complete ensemble of 12 NMR structures. See legend to **Figure 4** for base color-coding.

## BOX 5 | AUTOMATIC GENERATION OF STRUCTURED WATERS AROUND BASE PAIRS AND DINUCLEOTIDE STEPS IN NUCLEOSOMAL DNA

When performed with the '-w' option, *find\_pair* identifies the waters in contact with each base pair and *analyze* identifies the waters in contact with a dinucleotide step. Although the two programs select, by default, the waters, that is, the oxygen atoms of the waters, within 3.2 Å of any nitrogen or oxygen atom on a nucleotide pair or dimer fragment, the criteria are easily modified through a parameter file ('misc\_3dna.par'). The images shown here—the pattern of waters (denoted by red dots) associated with all the A-T (**Fig. 8a**) and all the G-C (**Fig. 8b**) nucleotide pairs, and the waters in contact with the first AT/AT dinucleotide step (**Fig. 8c**) in the currently best-resolved (1.9 Å) nucleosome core-particle structure<sup>71</sup>—are automatically generated using a Perl script that combines various 3DNA programs with PyMOL to generate the ray-traced images. Note that the A-T and G-C base pairs in the composite images are naturally overlapped on the middle base-pair reference frames, generated directly from *find\_pair*. This method forms the basis of the Solvation Web Service nucleic-acid solvation server<sup>35</sup>. Here, note the distinctive build-up of water near specific base, sugar and phosphate atoms.

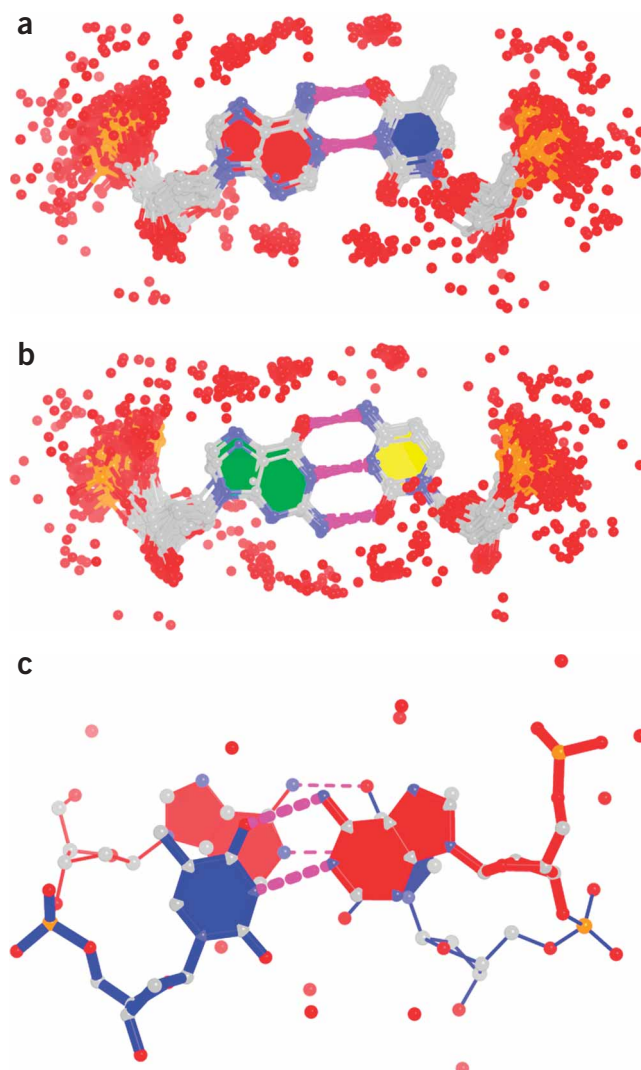
The stacking diagram that is also produced (**Fig. 4b**) clearly shows the inversion of the two base pairs, with their opposing connections to the sugar rings and differently exposed base-pair edges. The B-Z junction step has a twist angle of around  $-17^\circ$ , as well as noticeable shift (1.26 Å) and slide (0.80 Å) (given in the file '2acj.out' from Step 1C(iv)). The stacking diagram also reveals the significant overlap of bases on the same-strand (4.88 Å<sup>2</sup> for C10-T12 and 7.56 Å<sup>2</sup> for A25-G27, including exocyclic atoms) and the total absence of overlap of bases on complementary strands (i.e., cross-strand overlap) at the junction.

### (D) Automatic identification of double-helical regions in a DNA-RNA junction

The procedure in Step 1D illustrates how 3DNA can automatically identify the three helical regions in the DNA-RNA junction (**Fig. 5**). Helix 1, highlighted by the red cylinder, is made up of one continuous RNA chain along one strand (chain A in red) and two DNA chains along the other strand (chains B and D in the order of their 5'-3' directions). Helix 3, highlighted by the green cylinder, also contains one RNA chain along one strand (chain C in green) and two DNA chains along the other strand (chains D and B in order of their 5'-3' directions). Helix 2 consists primarily of DNA (chains B and D) with a single residue (A101) from the RNA chain inserted in one strand and another residue (A1) of the RNA along the other strand. Unlike helices 1 and 3, helix 2 is irregular and not represented by a cylinder. Thus, the backbones are not continuous along each helix. The chosen view allows for easy visualization and protractor measurement of the overall bending angle between the two relatively straight helices (the red horizontal helix h1 and the green vertical helix h3). This intuitive approach, initially developed to quantify the bending angles of a series of catabolic activator protein-DNA complexes<sup>67</sup>, yields sensible results when other methods fail.

### (E) Identification of higher-order base associations in ribosomal RNA

The images of the two base pentaplets in **Figure 6**, extracted from the large 50S ribosomal subunit and visualized with Step 1E, reveal both the canonical hydrogen-bonding of the Watson-Crick base pairs within the complexes and the noncanonical hydrogen bonding of the RNA base atoms with



**Figure 8** | DNA hydration patterns automatically extracted and illustrated with 3DNA from the currently best-resolved nucleosome core-particle structure<sup>71</sup>. The depicted waters (red dots) lie within 3.2 Å of the nucleotide oxygen and nitrogen atoms (a) in 84 of the 87 A-T pairs (three pairs are distorted) and (b) in all 60 G-C pairs and (c) in the first AT/AT dinucleotide step. Composite images (a,b) are generated by superposition of the middle frames of each base pair. See legend to **Figure 4** for base color-coding.

## BOX 6 | FIBER-DIFFRACTION MODELS

The *fiber* program within 3DNA provides handy access to 55 fiber models of DNA and RNA helices in various polymorphic forms (e.g., A-, B-, C-DNA and so on) and stoichiometries (e.g., single-stranded, double-stranded, triplexes, quadruplexes and DNA–RNA hybrids). The program generates structure files in standard PDB format, with the ‘-xml’ option allowing for structural output in the new Protein Data Bank Markup Language<sup>72</sup>, a format especially useful for very large structures that exceed the five-digit atom serial number limit and/or the fixed (f8.3) Cartesian-coordinate formatting limit of PDB files. To the best of our knowledge, the collection of fiber models in 3DNA is the most comprehensive of its kind. In preparing this set of fiber models, we have taken great care to ensure the accuracy and consistency of the models. For completeness and user verification, 3DNA includes, in addition to 3DNA-processed files, the original coordinates collected from the literature. Whereas the 3DNA-processed repeating unit is necessary to construct full, atomic-level representations of the models, the structurally equivalent PDB files of the repeating units, expressed in standard base reference frame, prove useful as building blocks when a user wishes to generate a perturbed base-only fiber model, such as an overstretched duplex with variable rise at successive steps.

the O2' atom of ribose and the exocyclic O2P phosphate oxygen, for example, O2' in the A2010–G2013 and G1971–U1972 (Fig. 6a) and the U2278–G2471 (Fig. 6b) pairs and O2P in the G1971–U1972 (Fig. 6a) and the U2278–G2471 and G2630–A2633 (Fig. 6b) pairs. The selected viewpoints highlight the very different types of interactions of adenine with the minor-groove edges of the G–C pairs in the two pentaplets, that is, distinct types of A-minor motifs<sup>68</sup>. Whereas A2010 approaches G2013 (Fig. 6a) with a different face exposed to the viewer (i.e., the positive z axes of the two purines point in opposite directions<sup>24</sup>), A2633 and G2471 (Fig. 6b) reveal the same face, with the two z axes pointing away from the viewer into the plane of the paper. Moreover, each of the adenines serves as a bridge that connects the two components in each pentaplet, concomitantly interacting with the canonical G–C pair through its minor-groove side and another base via its major-groove edge: with U1972 in a G–U dinucleotide platform<sup>69</sup> (Fig. 6a) or G2630 (Fig. 6b). It is also worth noting that the G1971–U1972 platform is stabilized not only by the well-characterized G(N2) · · · U(O4) H-bond interaction, but also by a little-noticed G(O2') · · · U(O2P) sugar-phosphate backbone interaction (Fig. 6a). Examination of the 50S large ribosomal unit (1JJ2) alone reveals ten such double H-bonded G–U platforms, far more occurrences than those registered by any other dinucleotide platform (including A–A) in this structure. Apparently, the G–U platform is more stable than other platforms with only a single base–base H-bond interaction. We are currently investigating this overrepresented G–U dinucleotide platform in other RNA structures.

### Other applications

3DNA contains more programs and features than can be fully documented here. We conclude with a brief mention of three other components of the software package that should be of general interest. Box 4 (see also Fig. 7) presents an example showing the automatic generation of schematic images of the ensemble of stem-hairpin structures found by Zoll *et al.*<sup>70</sup> from NMR studies of the 3'-UTR Y stem from the poliovirus RNA genome. Box 5 (see also Fig. 8) illustrates the procedure within 3DNA used to identify and visualize the structured water molecules found in high-resolution crystal structures, here the waters in contact with the base pairs and backbone of the DNA in the currently best-resolved nucleosome core-particle structure<sup>71</sup>. Finally, Box 6 describes the large collection of fiber-diffraction models that can be conveniently constructed with 3DNA.

Note: Supplementary information is available on the Nature Protocols website.

**ACKNOWLEDGMENTS** This work has been generously supported by the U.S. Public Health Service under research grant GM20861. We are grateful to Mauricio Esguerra and Guohui Zheng for their careful reading of the manuscript and independent validation of the protocols presented in this manuscript and to the users of 3DNA for using the software to address real-world problems, and communicating the difficulties that they encounter. Positive interactions with the users have been the driving force behind the development and improvement of 3DNA. We also thank the editor and the anonymous reviewers whose comments helped to clarify the presentation of the protocols.

Published online at <http://www.natureprotocols.com/>  
Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Tolstorukov, M.Y., Colasanti, A.V., McCandlish, D., Olson, W.K. & Zhurkin, V.B. A novel ‘roll-and-slide’ mechanism of DNA folding in chromatin. Implications for nucleosome positioning. *J. Mol. Biol.* **371**, 725–738 (2007).
2. Luscombe, N.M., Laskowski, R.A. & Thornton, J.M. Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Res.* **29**, 2860–2874 (2001).
3. Lu, X.-J., Shakked, Z. & Olson, W.K. A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819–840 (2000).
4. Tinoco, I. Jr. & Bustamante, C. How RNA folds. *J. Mol. Biol.* **293**, 271–281 (1999).

5. Auffinger, P. & Westhof, E. Rules governing the orientation of the 2'-hydroxyl group in RNA. *J. Mol. Biol.* **274**, 54–63 (1997).
6. Leontis, N.B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512 (2001).
7. Dickerson, R.E. *et al.* Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* **208**, 787–791 (1989).
8. Lu, X.-J. & Olson, W.K. Resolving the discrepancies among nucleic acid conformational analyses. *J. Mol. Biol.* **285**, 1563–1575 (1999).
9. Olson, W.K., Gorin, A.A., Lu, X.-J., Hock, L.M. & Zhurkin, V.B. DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proc. Natl. Acad. Sci. USA* **95**, 11163–11168 (1998).
10. Olson, W.K., Colasanti, A.V., Lu, X.-J. & Zhurkin, V.B. Physico-chemical properties of nucleic acids: character and recognition of Watson–Crick base pairs. In *Wiley Encyclopedia of Chemical Biology*. doi: 10.1002/9780470048672.wecb452 (John Wiley & Sons, Hoboken, NJ, USA, 2008).
11. Olson, W.K. *et al.* A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313**, 229–237 (2001).
12. Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. The complete atomic structure of the large ribosomal subunit at 2.4 Ångstrom resolution. *Science* **289**, 905–920 (2000).
13. Mandel-Gutfreund, Y., Margalit, H., Jernigan, R.L. & Zhurkin, V.B. A role for C · · · O interactions in protein–DNA recognition. *J. Mol. Biol.* **277**, 1129–1140 (1998).
14. Treger, M. & Westhof, E. Statistical analysis of atomic contacts at RNA–protein interfaces. *J. Mol. Recognit.* **14**, 199–214 (2001).



15. Suzuki, M. DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. *Proc. Natl. Acad. Sci. USA* **91**, 12357–12363 (1994).
16. Tatano, M. *et al.* DNA recognition by  $\beta$ -sheets. *Biopolymers* **44**, 335–359 (1997).
17. Jones, S., van Heyningen, P., Berman, H.M. & Thornton, J.M. Protein-DNA interactions: a structural analysis. *J. Mol. Biol.* **287**, 877–896 (1999).
18. Nadassy, K., Wodak, S.J. & Janin, J. Structural features of protein-nucleic acid recognition sites. *Biochemistry* **38**, 1999–2017 (1999).
19. Luscombe, N.M., Laskowski, R.A. & Thornton, J.M. NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res.* **25**, 4940–4945 (1997).
20. Woda, J., Schneider, B., Patel, K., Mistry, K. & Berman, H.M. An analysis of the relationship between hydration and protein-DNA interactions. *Biophys. J.* **75**, 2170–2177 (1998).
21. Kono, H. & Sarai, A. Structure-based prediction of DNA target sites by regulatory proteins. *Protein-Struct. Func. Genet.* **35**, 114–131 (1999).
22. Pichierrri, F., Aida, M., Gromiha, M.M. & Sarai, A. Free-energy maps of base-amino acid interactions for DNA-protein recognition. *J. Am. Chem. Soc.* **121**, 6152–6157 (1999).
23. Ge, W., Schneider, B. & Olson, W.K. Knowledge-based elastic potentials for docking drugs or proteins with nucleic acids. *Biophys. J.* **88**, 1166–1190 (2005).
24. Lu, X.-J. & Olson, W.K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121 (2003).
25. Zhurkin, V.B., Lysov, Y.P. & Ivanov, V.I. Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Res.* **6**, 1081–1096 (1979).
26. Bolshoy, A., McNamara, P., Harrington, R.E. & Trifonov, E.N. Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. USA* **88**, 2312–2316 (1991).
27. El Hassan, M.A. & Calladine, C.R. The assessment of the geometry of dinucleotide steps in double-helical DNA: a new local calculation scheme. *J. Mol. Biol.* **251**, 648–664 (1995).
28. Lu, X.-J., El Hassan, M.A. & Hunter, C.A. Structure and conformation of helical nucleic acids: analysis program (SCHNAAP). *J. Mol. Biol.* **273**, 668–680 (1997).
29. Lu, X.-J., El Hassan, M.A. & Hunter, C.A. Structure and conformation of helical nucleic acids: rebuilding program (SCHNAP). *J. Mol. Biol.* **273**, 681–691 (1997).
30. Lu, X.-J., Babcock, M.S. & Olson, W.K. Mathematical overview of nucleic acid analysis programs. *J. Biomol. Struct. Dynam.* **16**, 833–843 (1999).
31. Berman, H.M. *et al.* The Nucleic Acid Database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751–759 (1992).
32. Sühnel, J. Views of RNA on the World Wide Web. *Trends in Genetics* **13**, 206–207 (1997).
33. Merritt, E.A. & Bacon, D.J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
34. Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
35. Auffinger, P. & Hashem, Y. SwS: a solvation web service for nucleic acids. *Bioinformatics* **23**, 1035–1037 (2007).
36. Dixit, S.B. & Beveridge, D.L. Structural bioinformatics of DNA: a web-based tool for the analysis of molecular dynamics results and structure prediction. *Bioinformatics* **22**, 1007–1009 (2006).
37. Dror, O., Nussinov, R. & Wolfson, H. The ARTS web server for aligning RNA tertiary structures. *Nucleic Acids Res.* **34**, W412–W415 (2006).
38. van Dijk, M., van Dijk, A.D.J., Hsu, V., Boelens, R. & Bonvin, A.M.J.J. Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. *Nucleic Acids Res.* **34**, 3317–3325 (2006).
39. Banerjee, A. *et al.* Feature extraction using molecular planes for fuzzy relational clustering of a flexible dopamine reuptake inhibitor. *J. Chem. Inf. Model.* **47**, 2216–2227 (2007).
40. Babcock, M.S., Pednault, E.P.D. & Olson, W.K. Nucleic acid structure analysis: mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.* **237**, 125–156 (1994).
41. Clowney, L., Jain, S.C., Srinivasan, A.R., Westbrook, J., Olson, W.K. & Berman, H.M. Geometric parameters in nucleic acids: nitrogenous bases. *J. Am. Chem. Soc.* **118**, 509–518 (1996).
42. Dickerson, R.E. DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* **26**, 1906–1926 (1998).
43. Gorin, A.A., Zhurkin, V.B. & Olson, W.K. B-DNA twisting correlates with base pair morphology. *J. Mol. Biol.* **247**, 34–48 (1995).
44. Lavery, R. & Sklenar, H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6**, 63–91 (1988).
45. Lavery, R. & Sklenar, H. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dynam.* **6**, 655–667 (1989).
46. Macke, T. & Case, D.A. Modeling unusual nucleic acid structures. In *Molecular Modeling of Nucleic Acids* (eds. Leontis, N.B. & SantaLucia J. Jr.) 379–393 (American Chemical Society, Washington, DC, 1998).
47. Tung, C.-S. & Carter, E.S. II Nucleic acid modeling tool (NAMOT): an interactive graphic tool for modeling nucleic acid structures. *Bioinformatics* **10**, 427–433 (1994).
48. Major, F. Building three-dimensional ribonucleic acid structures. *IEEE Comput. Sci. Eng.* **5**, 44–53 (2003).
49. Sayle, R.A. & Milnerwhite, E.J. RasMol: biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374–376 (1995).
50. Massire, C., Gaspin, C. & Westhof, E. DRAWNA: a program for drawing schematic views of nucleic acids. *J. Mol. Graph.* **12**, 201–206 (1994).
51. Huang, C.C., Couch, G.S., Pettersen, E.F. & Ferrin, T.E. Chimera: an extensible molecular modeling application constructed using standard components. *Pac. Symp. Biocomput.* **1**, 724 (1996).
52. Trifonov, E.N. & Sussman, J.L. The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc. Natl. Acad. Sci. USA* **77**, 3816–3820 (1980).
53. Hagerman, P.J. Sequence dependence of the curvature of DNA: a test of the phasing hypothesis. *Biochemistry* **24**, 7033–7037 (1985).
54. Koo, H.-S., Wu, H.-M. & Crothers, D.M. DNA bending at adenine thymine tracts. *Nature* **308**, 501–506 (1986).
55. Hagerman, P.J. Sequence-directed curvature of DNA. *Nature* **321**, 449–450 (1986).
56. Ulanovsky, L.E. & Trifonov, E.N. Estimation of wedge components in curved DNA. *Nature* **326**, 720–722 (1987).
57. Schultz, S.C., Shields, G.C. & Steitz, T.A. Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. *Science* **253**, 1001–1007 (1991).
58. Parkinson, G., Wilson, C., Gunasekera, Ebright, Y.W., Ebright, R.E. & Berman, H.M. Structure of the CAP-DNA complex at 2.5 Å resolution: a complete picture of the protein-DNA interface. *J. Mol. Biol.* **260**, 395–408 (1996).
59. Richmond, T.J. & Davey, C.A. The structure of DNA in the nucleosome core. *Nature* **423**, 145–150 (2003).
60. Calladine, C.R., Drew, H.R., Luisi, B.F. & Travers, A.A. *Understanding DNA: the Molecule and How It Works* Chapter 4 (Elsevier Academic Press, San Diego, CA, 2004).
61. Olson, W.K., Srinivasan, A.R., Marky, N.L. & Balaji, V.N. Theoretical probes of DNA conformation examining the B leads to Z conformational transition. *Cold Spring Harb. Symp. Quant. Biol.* **47**, 229–241 (1983).
62. Harvey, S.C. DNA structural dynamics: longitudinal breathing as a possible mechanism for the B->Z transition. *Nucleic Acids Res.* **11**, 4867–4878 (1983).
63. Ha, S.C., Lowenhaupt, K., Rich, A., Kim, Y.-G. & Kim, K.K. Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437**, 1183–1186 (2005).
64. Nowakowski, J., Shim, P.J., Stout, C.D. & Joyce, G.F. Alternative conformations of a nucleic acid four-way junction. *J. Mol. Biol.* **300**, 93–102 (2000).
65. Klein, D.J., Schmeing, T.M., Moore, P.B. & Steitz, T.A. The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221 (2001).
66. Matsumoto, A. & Olson, W.K. Sequence-dependent motions of DNA: a normal mode analysis at the base-pair level. *Biophys. J.* **83**, 22–41 (2002).
67. Chen, S., Vojtechovsky, J., Parkinson, G.N., Ebright, R.H. & Berman, H.M. Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: DNA binding specificity based on energetics of DNA kinking. *J. Mol. Biol.* **314**, 63–74 (2001).
68. Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B. & Steitz, T.A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. USA* **98**, 4899–4903 (2001).
69. Wimberly, B.T., Guymon, R., McCutcheon, J.P., White, S.W. & Ramakrishnan, V. A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell* **97**, 491–502 (1999).
70. Zoll, J., Tessari, M., Van Kuppeveld, F.J.M., Melchers, W.J.G. & Heus, H.A. Breaking pseudo-twofold symmetry in the poliovirus 3'-UTR Y-stem by restoring Watson-Crick base pairs. *RNA* **13**, 781–792 (2007).
71. Davey, C.A., Sargent, D.F., Luger, K., Mäder, A.W. & Richmond, T.J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).
72. Westbrook, J., Ito, N., Nakamura, H., Henrick, K. & Berman, H.M. PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics* **7**, 988–992 (2005).