DNA Structures

DOI: 10.1002/anie.201003647

# A Geometric Approach to the Crystallographic Solution of Nonconventional DNA Structures: Helical Superstructures of d(CGATAT)**

*Iñaki Martínez de Ilarduya, Daniela De Luchi, Juan A. Subirana, J. Lourdes Campos,* and Isabel Usón**

*Dedicated to Professor Herbert W. Roesky on the occasion of his 75th birthday*

Oligonucleotides have been used extensively to build nanostructures[1] and nanodevices.[2] Mostly, rather long oligonucleotides (10–100 bases) are used to form either flat tiles or closed objects based on standard Watson–Crick base pairs.[3] Only recently, self-assembled three-dimensional DNA lattices have been described.[4] Such structures feature large cavities, allowing incorporation of globular shaped molecules. Also, short oligonucleotides (2–12 bases) may assemble into intricate lattices, such as cubes[5] and other complex structures,[6] containing large voids. X-ray crystallography provides the indispensable three-dimensional view into the atomic structure of such nanomaterials, but their crystals usually diffract far from atomic resolution, and thus their structures cannot be solved by direct methods.[7]

Herein, we present a new geometrical approach to solve nonconventional DNA structures and its application to the solution of the superstructure of new topology formed by d(CGATAT). Such unprecedented structures could result in materials with new properties, and their characterization should not be hindered for lack of a suitable phasing method.

As it is well known from the pioneering work on the structure of the DNA double helix using fiber diffraction,[8] in contrast to protein crystallography, meaningful information can be derived already from the diffraction pattern. Previous knowledge leads to the expectation that DNA forms base-paired, double-stranded helices in A, B, or Z geometry. Such helical moieties tend to stack on piles or to lean their ends on the grooves of other helices. Other motifs may play a role: quadruplexes,[9] three-way and Holliday-junctions,[10] or looping out unpaired bases[11] have been described in the structures of oligonucleotides and their complexes. Major base-stacking directions can be identified from the diffraction images by the strong Bragg reflections at 3.3 Å spacing and fiber streaks. Thus, analysis of the diffraction data fixes the preferred orientation of piled base pairs (Figure S1a in the Supporting Information). The unit-cell geometry and the symmetry, along with the estimation of the solvent content from the atomic volumes, allow one to predict whether they fit a simple packing of regular helices or a distortion is required to build a three-dimensional structure. Figure S1b illustrates the relationship between the dimensions of a hexagonal projection and the requirements on the helical radii. Thus, examination of the geometrical parameters can be exploited to set up structural hypotheses as to the building blocks present and their packing, to be confirmed or discarded through molecular replacement or refinement of the models.

To identify such models, we automated the analysis of the packing of all DNA structures deposited with the Protein or Nucleic Acid Databases (PDB/NDB). Our program SUBIX (Figure 1) allows one to establish the geometrical requirements of different projections and to classify DNA materials according to their building blocks, thus identifying or assembling the best candidates to be used alone or in combination

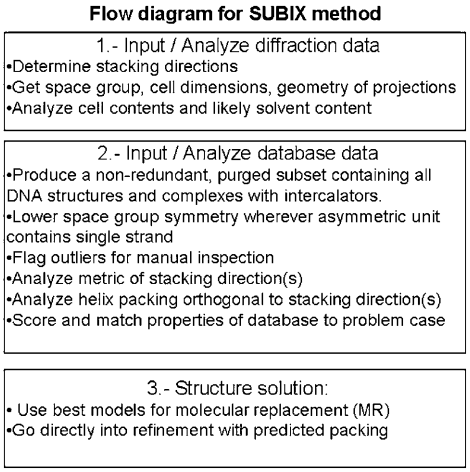


***Figure 1.*** Flow diagram illustrating the elements of the geometric approach to structure solution. A SUBIX test version is available from the authors upon request. The program will be distributed freely to academic users once testing and debugging is complete.

[*] I. Martínez de Ilarduya,[+] Prof. Dr. I. Usón
Institucio Catalana de Recerca i Estudis Avançats (ICREA) at Instituto de Biología Molecular de Barcelona IBMB-CSIC, Barcelona Science Park, Baldiri Reixach, 13, Barcelona 08028 (Spain)
E-mail: uson@ibmb.csic.es
Dr. D. De Luchi,[+] Prof. Dr. J. A. Subirana, Dr. J. L. Campos
Departament d'Enginyeria Química, Universitat Politècnica de Catalunya, Diagonal 647, Barcelona 08028 (Spain)
E-mail: lourdes.campos@upc.edu
[+] These authors contributed equally to this work.

[**] This work was supported by grants BFU 2009-10380 and BIO 2009-10576, MICINN, Spain. Data collection was supported by the ESRF (BM16) and the EU.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201003647.

as search fragments for molecular replacement. It should be stressed that, as a fundamental difference to standard protein molecular replacement, this process is performed in a purely geometrical way, by considering the length of the sequence but disregarding its composition. The geometric rational can be applied to any space group, but it will be illustrated in the case leading to the structure solution of a previously unknown DNA hexagonal structure, which could not be solved by molecular replacement with an ATAT B-DNA fragment.

Crystals of the hexanucleotide CGATAT belong to the space groups $P6_222$ or $P6_422$ ($a=b=38.33$, $c=56.82$ Å) and diffract well for DNA: 2.6 Å resolution despite the rather high solvent content of 64%. The diffraction pattern shows evidence of stacking approximately along the $Z$ axis (Figure S1a). Database analysis (summarized in Table S1) shows that typically DNA helices in trigonal or hexagonal space groups that pack aligned with the $Z$ direction are arranged in contiguous or intercrossing columns. Interestingly, no occurrences showing this disposition have been reported in $P6_222$ or $P6_422$ space groups or their hexagonal subgroups ($P6_2$, $P6_4$) or in $P3$, $P3_1$, $P3_121$, $P3_112$, and $P6_1$. For the case of piles of helices, the typical $a=b$ dimensions cluster around 27 Å, in space groups $P3_2$, $P3_221$, $P3_212$, and $P6_122$ or around 40 Å in $R3$ and $R32$. Both cases (effectively equivalent owing to the centering) display rather compact packing, with a solvent content around 45%, which is very far from the case in our structure. Space group $P6$, in contrast, leads to packing with a higher solvent content (ca. 50%) but increased unit cell constants $a=b=53$ Å. $P6_5$ and $P6_522$ space groups contain Z-DNA structures with $a$ and $b$ values of 36 and 18 Å, respectively. $P6_3$ and $P6_322$ gather cases with a distorted, smaller helix radius as a result of the autointercalation of guanine residues or the presence of bisintercalators such as triostin and echinomycin, leading to a less twisted helix and $a$ and $b$ unit-cell constants of around 37 Å for the former and around 40 Å for the latter, as well as a significant elongation in the $c$ axis. For an average helix, the stacking repeat must correspond approximately to a multiple of 3.3 times the number of stacked base pairs and intercalators, multiplied by the number of copies required by the symmetry of the space group. In the case of CGATAT, 56.8 for six nucleotides fits a slightly short stacking parameter of 3.2 Å. Finally, the model with the best geometrical score was PDB entry 231D. This structure contains the sequence CGATCG and a bulky porphyrin group, straining the helix radius to an increased value. It packs in the space group $P6_122$ with unit-cell values of $a=b=39.49$ and $c=56.15$, and an estimated solvent content of 51%. A double-stranded model derived from this structure without the porphyrin intercalators was used to solve the structure in the space group $P6_2$ with the program MOLREP.[12]

Figure 2 illustrates the structure solution. Discrimination of the correct solution is not straightforward, as the search model (Figure 2a) is extremely different and, typically, all solutions with correct stacking have similarly high figures of merit. Packing examination (Figure 2b) is a good indication but cannot be too stringent. Refinement of the correct solution with REFMAC5[13] not only produces sensible residual values, but, more importantly, the difference electron density distinctly reveals the presence of wrong and missing features in the start model (Figure 2c). The final $P6_222$ model refined to satisfactory values for comparable DNA structures ($R_{work}=0.243$, $R_{free}=0.278$) (Figure 2d).

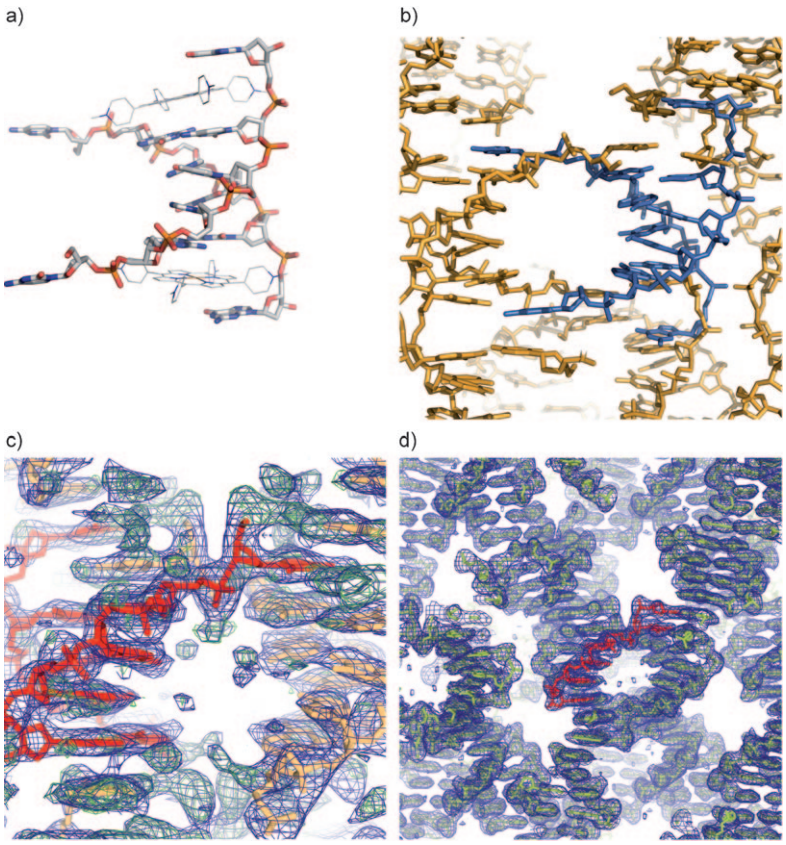


***Figure 2.*** Stages in the structure-solution process. a) Model used for molecular replacement; b) packing of the best solution obtained with CGATCG from PDB entry 231D; c) $\sigma_A$-weighted $2F_o-F_c$ (blue) and $F_o-F_c$ (green for positive) electron-density maps, contoured at 1σ and 2σ, respectively; d) packing of the final model and final $\sigma_A$-weighted $2F_o-F_c$ (blue) electron-density map contoured at 1σ. Figure prepared with Pymol molecular software package (Schrodinger llc.)

The central unit of the structure is a B-form duplex built by four A·T base pairs (Figure 3a). The terminal CG bases stick out and form a Z-form duplex with a neighbor molecule. Z-form duplexes from another group of molecules are stacked on the terminal base pairs of the AT helix (Figure 3b). The complex set of helical structures in the crystal (Figure 3c,d; see also the Supporting Information) contrasts with other helical superstructures.[14] In the related CGATATATATAT[15] and CGACGATCGT sequences,[16] the CG sticky ends lead to a continuous double-helical structure, whereas

  

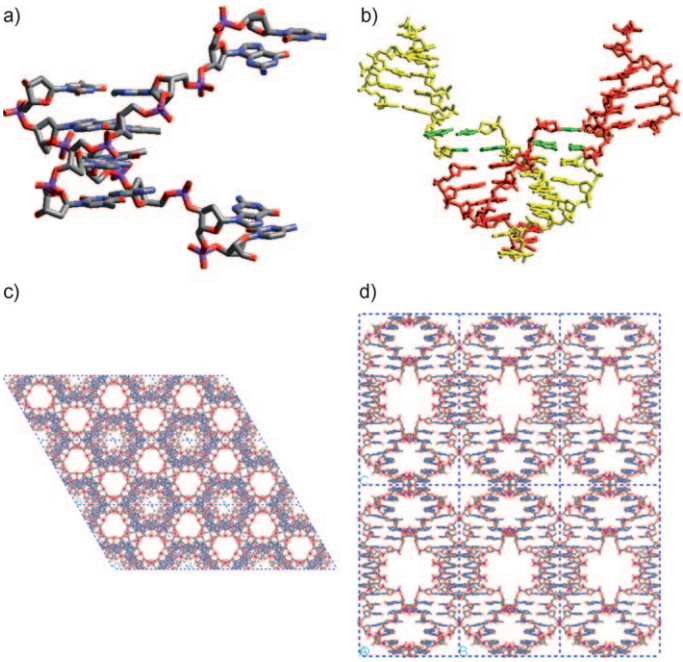


***Figure 3.*** a) The basic unit in the crystal: a Watson–Crick DNA duplex with sticky ends. b) Scheme of the interactions that give rise to higher-order structures. A group of two hydrogen-bonded duplexes is shown in yellow, another one in red. The Z-form C·G base pairs are shown in green. The latter are stacked onto a neighbor B-form duplex. A stereoview is given in the Supporting Information (Figure S1). c) Projection of 18 unit cells of the crystal onto the *ab* plane and d) onto the *bc* plane. Large cavities appear in different orientations in space.

CGATAT forms two independent helices. The shorter CG part of the molecule adopts the Z-form because its intrinsic preference for this form[17] drives the supramolecular arrangement.

This work reveals that complex helical superstructures can be designed and built from small molecules. As far as we know, the topological arrangement of a large number of helices crossing in space had not been previously described. We expect that other structures presenting such distortions have remained unsolved for lack of an appropriate phasing method. Thus, our approach should contribute to characterizing additional nonstandard cases and hence to develop further our analysis to the prediction of new DNA-based materials of nanotechnological interest.

## Experimental Section

Crystals were obtained by vapor diffusion techniques at 13 °C using 80 mM magnesium acetate, 4 mM spermine, pH 6.5, and PEG 4000 as precipitant. Data were collected on beamline BM16 at the ESRF and processed with HKL2000.[18] Graphical building was done with COOT.[19] The superstructures were analyzed with CERIUS-2 (Accelrys, Inc.). Coordinates and structure factors are deposited at the PDB with code 3HCL.

Received: June 15, 2010
Revised: August 2, 2010
Published online: September 16, 2010

**Keywords:** DNA · helical structures · oligonucleotides · structure elucidation · supramolecular chemistry

[1] N. C. Seeman, *Chem. Biol.* **2003**, *10*, 1151; a) U. Feldkamp, C. M. Niemeyer, *Angew. Chem.* **2006**, *118*, 1888; *Angew. Chem. Int. Ed.* **2006**, *45*, 1856; b) M. G. Svahn, M. Hasan, V. Sigot, J. J. Valle-Delgado, M. W. Rutland, K. E. Lundin, C. I. Smith, *Oligonucleotides* **2007**, *17*, 80; c) M. Brucale, G. Zuccheri, B. Samori, *Trends Biotechnol.* **2006**, *24*, 235; d) N. C. Seeman, *Methods Mol. Biol.* **2005**, *303*, 143; e) F. C. Simmel, *Angew. Chem.* **2008**, *120*, 5968; *Angew. Chem. Int. Ed.* **2008**, *47*, 5884; f) F. A. Aldaye, A. L. Palmer, H. F. Sleiman, *Science* **2008**, *321*, 1795.

[2] a) M. K. Beissenhirtz, I. Willner, *Org. Biomol. Chem.* **2006**, *4*, 3392; b) B. Yurke, A. J. Turberfield, A. P. Mills, Jr., F. C. Simmel, J. L. Neumann, *Nature* **2000**, *406*, 605; c) H. Yan, X. Zhang, Z. Shen, N. C. Seeman, *Nature* **2002**, *415*, 62; d) C. Mao, W. Sun, Z. Shen, N. C. Seeman, *Nature* **1999**, *397*, 144; e) B. Ding, N. C. Seeman, *Science* **2006**, *314*, 1583; f) C. Lin, Y. Liu, H. Yan, *Biochemistry* **2009**, *48*, 1663.

[3] a) F. A. Aldaye, H. F. Sleiman, *J. Am. Chem. Soc.* **2007**, *129*, 13376; b) W. Wang, Y. Yang, E. Cheng, M. Zhao, H. Meng, D. Liu, D. Zhou, *Chem. Commun.* **2009**, 824; c) J. H. Chen, N. C. Seeman, *Nature* **1991**, *350*, 631; d) R. Goodman, R. Berry, A. Turberfield, *Chem. Commun.* **2004**, 1372; e) R. P. Goodman, M. Heilemann, S. Doose, C. M. Erben, A. N. Kapanidis, A. J. Turberfield, *Nat. Nanotechnol.* **2008**, *3*, 93; f) C. Zhang, M. Su, Y. He, X. Zhao, P. A. Fang, A. E. Ribbe, W. Jiang, C. D. Mao, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10665; g) D. Bhatia, S. Mehtab, R. Krishnan, S. S. Indi, A. Basu, Y. Krishnan, *Angew. Chem.* **2009**, *121*, 4198; *Angew. Chem. Int. Ed.* **2009**, *48*, 4134.

[4] a) P. J. Paukstelis, J. Nowakowski, J. J. Birktoft, N. C. Seeman, *Chem. Biol.* **2004**, *11*, 1119; b) J. Zheng, J. J. Birktoft, Y. Chen, T. Wang, R. Sha, P. E. Constantinou, S. L. Ginell, C. Mao, N. C. Seeman, *Nature* **2009**, *461*, 74.

[5] N. Valls, I. Usón, C. Gouyette, J. A. Subirana, *J. Am. Chem. Soc.* **2004**, *126*, 7812.

[6] L. Malinina, L. Urpí, X. Salas, T. Huynh-Dinh, J. A. Subirana, *J. Mol. Biol.* **1994**, *243*, 484.

[7] G. M. Sheldrick, H. Hauptmann, C. Weeks, R. Miller, I. Usón in *Int. Tables for Crystallography F* (Eds.: E. Arnold, M. Rossmann), Kluwer Academic Publishers, Dordrecht, **2001**, p. 333.

[8] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737.

[9] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402.

[10] a) Y. Liu, S. C. West, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 937; b) A. Oleksi, A. G. Blanco, R. Boer, I. Usón, J. Aymami, A. Rodger, M. J. Hannon, M. Coll, *Angew. Chem.* **2006**, *118*, 1249; *Angew. Chem. Int. Ed.* **2006**, *45*, 1227.

[11] L. Joshua-Tor, D. Rabinovich, H. Hope, F. Frolow, E. Apella, J. L. Sussman, *Nature* **1988**, *334*, 82.

[12] A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, *30*, 1022.

[13] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. Sect. D* **1997**, *53*, 240.

[14] a) J. L. Campos, L. Urpí, T. Sanmartín, C. Gouyette, J. A. Subirana, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3663; b) J. A. Subirana, M. Creixell, R. Baldini, J. L. Campos, *Biomacromolecules* **2008**, *9*, 6.

[15] D. De Luchi, V. Tereshko, C. Gouyette, J. A. Subirana, *ChemBioChem* **2006**, *7*, 585.

[16] H. Qiu, J. C. Dewan, N. C. Seeman, *J. Mol. Biol.* **1997**, *267*, 881.

[17] P. S. Ho, M. J. Ellison, G. J. Quigley, A. Rich, *EMBO J.* **1986**, *5*, 2737.

[18] Z. Otwinowski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307.

[19] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D* **2004**, *60*, 2126.

  