

Crystal Structure of a Complex of DNA with One AT-Hook of HMGA1

Elsa Fonfría-Subirós¹, Francisco Acosta-Reyes¹, Núria Saperas¹, Joan Pous², Juan A. Subirana¹, J. Lourdes Campos^{1*}

1 Departament d'Enginyeria Química, Universitat Politècnica de Catalunya, Barcelona, Spain, **2** Plataforma Automatitzada de Cristal·lografia, Institut de Recerca Biomèdica de Barcelona, PCB-CSIC, Barcelona, Spain

Abstract

We present here for the first time the crystal structure of an AT-hook domain. We show the structure of an AT-hook of the ubiquitous nuclear protein HMGA1, combined with the oligonucleotide d(CGAATTAATTCG)₂, which has two potential AATT interacting groups. Interaction with only one of them is found. The structure presents analogies and significant differences with previous NMR studies: the AT-hook forms hydrogen bonds between main-chain NH groups and thymines in the minor groove, DNA is bent and the minor groove is widened.

Citation: Fonfría-Subirós E, Acosta-Reyes F, Saperas N, Pous J, Subirana JA, et al. (2012) Crystal Structure of a Complex of DNA with One AT-Hook of HMGA1. *PLoS ONE* 7(5): e37120. doi:10.1371/journal.pone.0037120

Editor: Emanuele Buratti, International Centre for Genetic Engineering and Biotechnology, Italy

Received: January 27, 2012; **Accepted:** April 13, 2012; **Published:** May 16, 2012

Copyright: © 2012 Fonfría-Subirós et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants BFU2009-10380 from the Spanish Ministerio de Ciencia e Innovación and 2009 SGR 1208 from the Generalitat de Catalunya. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lourdes.campos@upc.edu

Introduction

High mobility group A proteins (HMGA, formerly called HMG-I/Y) are intrinsically disordered non-histone chromosomal proteins characterized by containing three DNA-binding domains, called AT-hooks, which preferentially bind to the minor groove of short stretches of AT-rich DNA [1]. By binding to differently spaced AT-rich DNA regions and/or direct interaction with several transcription factors, HMGAs regulate the expression of numerous genes. In this way they influence many normal biological processes including growth, proliferation, differentiation and death [2–4]. Alterations or abnormal expression of HMGA proteins have also been related to several pathological processes and metabolic disorders, including obesity, type 2 diabetes mellitus and cancer [5–7]. Thus, HMGA proteins have been proposed as therapeutic drug targets by several authors [7,8]. The nature of the interaction was already determined by Reeves and Nissen [1], who used foot-printing methods. Further details on the interaction were later determined by NMR methods [9, 10]. From the structural point of view no results are available on the interaction of HMGA with DNA, as determined by crystallographic methods. Here we present the crystal structure of the complex of a DNA oligonucleotide with the third AT-hook (DBD3 in ref. 10) of the HMGA1 protein. Their sequence is given in Figure 1c.

As mentioned above, the HMGA1 protein contains three AT-hooks. All of them are formed by a conserved core sequence, Pro-Arg-Gly-Arg-Pro, flanked by positively charged amino acids [1]. The first AT-hook (DBD1) presents a significant difference, since the first Pro is substituted by Gly. The other two AT-hooks (DBD2 and DBD3) only differ in the nature of the basic amino acids present next to the conserved core. Thus, the affinity of short peptides (10–11 amino acids) for DNA is similar both for DBD2 and DBD3 [11]. On the other hand, in the whole protein, the

affinity of DBD2 for DNA is higher [1,10,11], which has been explained by the presence of additional basic amino acids in neighbor regions of the protein and which are missing in DBD3 [1]. In our studies we have used a peptide with nine amino acids identical to those found in DBD3 and similar to those found in DBD2. Thus, our results are relevant for the interaction with DNA of both core sequences of DBD2 and DBD3.

Results and Discussion

As shown in the Methods section our crystals contain four complexes which are very similar. The geometry of the duplex and the central sequence of the AT-hook (PRGRP) are practically identical in the four complexes. On the other hand the external AT-hook terminal amino acids vary in each of the four complexes present in our crystal. Some of them form crosslinks with neighbor duplexes. In Figure 1 we present one of the complexes (oligonucleotide chains C, D and peptide M). The DNA duplex has a clear bend, localized in the region of base pairs 5 and 6. It determines an angle of 24° between the lower five bases and the upper seven bases of the duplex. This local bending may contribute to the overall bending detected when the whole protein is associated with DNA [12]. The DNA duplex has two AATT regions. As shown in Figure 1, the minor groove in one of them is occupied by the AT-hook, whereas the other region shows a clear spine of hydration, typical for the AATT sequence [13,14]. The minor groove is wide in the region occupied by the AT-hook, with phosphate-phosphate distances in the range 12–13 Å. In the upper region the minor groove is rather narrow, with phosphate-phosphate distances about 9–9.5 Å. Bending of the DNA is probably due to the presence of the AT-hook, which distorts the minor groove. Packing interactions between duplexes may also have an influence [15].

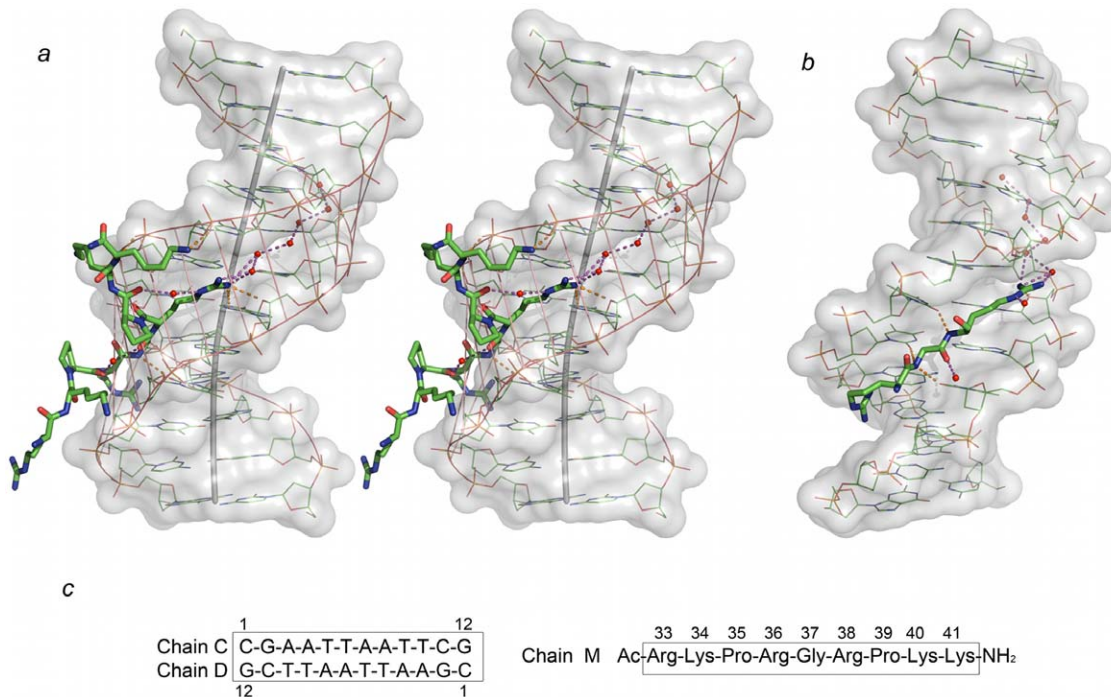


Figure 1. Stereoviews and composition of the AT-hook/DNA complex. **a** Stereoview of one AT-hook/DNA complex. The DNA is shown as a partially transparent object, drawn with Pymol. Its bent helical axis is shown in darker grey (calculated with CURVES+, available at http://gbio-pbil.ibcp.fr/Curves_plus/). Two water molecules which contribute to stabilize the structure are also represented, as well as waters in the spine of hydration. Virtual bonds between phosphates are indicated by thin red lines: they demonstrate a much wider minor groove in the region where the AT-hook is inserted. **b** In this view only the elongated central region of the AT-hook (Arg-Gly-Arg) is shown. Hydrogen bonds between main chain peptide NH atoms (Gly37 and Arg38) and thymine oxygen atoms are also indicated. **c** Numbering of the oligonucleotide and peptide residues used in this work. In order to facilitate comparison we have used the same numbering as reported for the DBD3 hook in the NMR structure [10]. doi:10.1371/journal.pone.0037120.g001

The dodecamer we have used has eight consecutive A·T base pairs, so that the AT-hook might be positioned in several places. However it is clearly associated with one of the AATT regions and occupies fully its minor groove. Interactions of the guanidinium groups of both arginines 36 and 38 extend further into the neighboring base pairs. This fact excludes the possibility of finding two AT-hooks associated with our dodecamer sequence. The

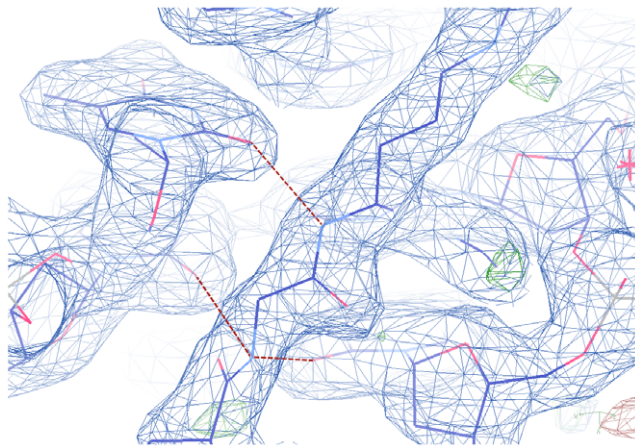


Figure 2. Electron density map ($2F_o-F_c$ at 1σ level) of a segment of the AT-hook in the minor groove of DNA. Hydrogen bonds between main chain peptide NH atoms (Gly37 and Arg38) and thymine oxygen atoms are indicated as red dashed lines. doi:10.1371/journal.pone.0037120.g002

position we have found for the AT-hook also confirms previous studies [16,17], which indicated that the AATT sequence is a favored site of interaction, whereas the TTAA sequence in our



Figure 3. In this figure the DNA duplex is shown as space filling atoms with their van der Waals radii. The strong association of the inner PRGRP sequence with the duplex is clearly visible. The terminal basic amino acids (34, 40 and 41) interact with phosphates in the same duplex, with the exception of Arg33 which does not show any apparent interactions. doi:10.1371/journal.pone.0037120.g003

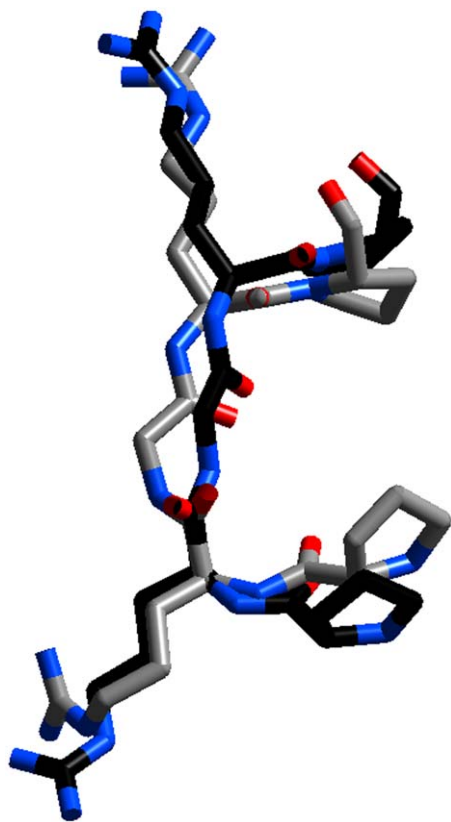


Figure 4. Comparison of the structures of the central PRGRP sequence of the AT-hook obtained by either X-ray diffraction (in grey, this work) or NMR [10]. The overall conformation is similar, but the X-ray structure has significant changes in the position of the main chain NH groups, which allow the formation of hydrogen bonds with minor groove thymine oxygen atoms of DNA. This figure has been prepared with Cerius2.
doi:10.1371/journal.pone.0037120.g004

dodecamer does not appear to be appropriate, according to the previous studies mentioned. Optimal van der Waals interactions of the AT-hook with adenines appear to be important for such specificity [10].

The electron density map of the central region of the AT-hook is shown in Figure 2. It is clear that the extended Arg-Gly-Arg chain is tightly bound to the minor groove, but individual atoms are not visible. However the protruding carbonyl oxygen atoms of the main chain are clearly visible and determine the overall geometry of the amino acids. A clear hydrogen bond of the main chain NH group of Arg-38 is formed with one thymine oxygen. An additional weaker bifurcated hydrogen bond is formed by the NH group of Gly and two thymine oxygen atoms from the opposite strands of the duplex, as shown in Figures 1b and 2. Another view of the complex is presented in Figure 3. The guanidinium group of Arg 38 is clearly apparent in the electron density map and forms multiple hydrogen bonds, with one thymine, one sugar oxygen and the first water in the neighbor spine of hydration. On the other hand the guanidinium group of Arg 36 appears only as a small blob in the electron density map and cannot be accurately positioned. It has multiple possible interactions with neighboring bases and it may have several alternative orientations. An additional stabilizing factor of the AT-hook is the presence of two water bridges, shown in Figure 1a. These water molecules form bridges between the carbonyls of Arg 38 and Pro 39 and

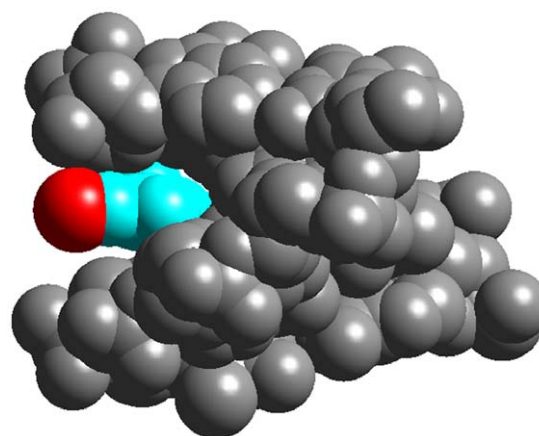


Figure 5. A model of the effect of methylation on Arg38 in the DBD3 AT-hook. Only the guanidinium group of Arg38 (in blue) is visible in this orientation. DNA is shown in grey. In this model the hydrogen bonds between the arginine guanidinium group and the minor groove of DNA are maintained, whereas the methyl group (in red) has an external position. In the side facing DNA there is no room for the methyl group, without disturbing the guanidinium-DNA hydrogen bonds.
doi:10.1371/journal.pone.0037120.g005

between Arg 36 and Pro 35. We also enclose a video (Video S1) as supporting information.

A comparison with the NMR structure [10] is shown in Figure 4. The overall conformation is rather similar, but there are two significant differences. In particular the conformation of the chain next to the central glycine is different. Our X-ray structure enters more deeply into the minor groove and has a different orientation of the main chain NH groups, which form hydrogen bonds with

Table 1. Crystal data and refinement statistics.

Wavelength (Å)	0.9795
Space group	P21
Unit-cell parameters (Å, °)	a = 49.143, b = 66.855, c = 50.052, $\alpha = \gamma = 90$ $\beta = 111.4$
Resolution range (Å)	40.00–2.27 (2.31–2.27)
Unique reflections	13163 (994)
Free R-factor reflections	740 (47)
Completeness (%)	98.73 (98.2)
Average Redundancy	7.5 (7.3)
$[I/\sigma(I)]$	2.2 (1.6)
R_{merge}	0.166 (0.346)
Contents of asymmetric unit	4 DNA duplexes, 4 peptide molecules, 59 water molecules. Total: 2289 non-H atoms.
R_{work}^a	0.225 (0.321)
R_{free}^b	0.271 (0.387)
Mean B factor (Å ²)	36.36
R.m.s.d. bonds (Å)	0.006
R.m.s.d. angles (°)	1.062

Values in parentheses are for the last shell.

^a $\sum_{hkl} |F_0(hkl) - KF_c(hkl)| / \sum_{hkl} F_0(hkl)$.

^bR factor of reflections used for cross validation in the refinement.

doi:10.1371/journal.pone.0037120.t001

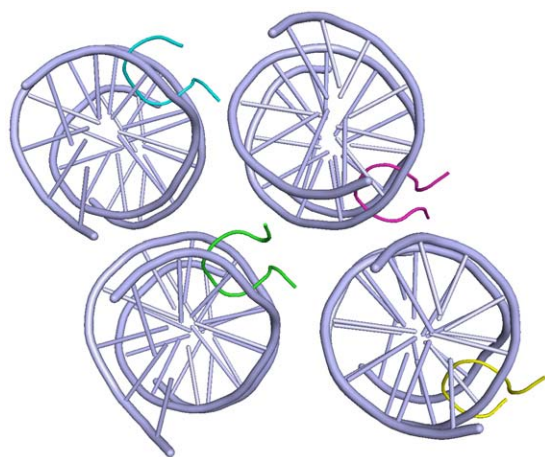


Figure 6. Simplified view of the asymmetric unit in the crystal.

The two complexes at right are in the up orientation, as shown in Figure 1 in the main text. The two complexes at left are in the opposite orientation, with the peptide running down. In the crystal each complex generates a continuous column of identical complexes related by a screw axis. A detailed analysis of the packing interactions will be reported elsewhere.

doi:10.1371/journal.pone.0037120.g006

thymines in the DNA. The presence of such hydrogen bonds was postulated in the pioneering work of Reeves and Nissen [1]. An additional difference is the position of the guanidinium group of Arg 36, which in the X-ray structure is not uniquely positioned. However the main lesson of our work is that the AT-hook can bind to a strongly distorted DNA molecule, whereas most models of interaction assume that the B form of DNA remains practically unchanged upon binding [11,18].

As a final point we may comment on what structural effect might have the methylation of arginine in AT-hooks. It is known that tumour cells contain methyl-arginine in different positions of HMGA1, as reviewed by Zhang and Wang [19]. The most common site of methylation occurs in the first AT-hook [20], which has a different sequence than the DBD3 we have studied. Methylation has no direct influence on the formation of hydrogen bonds by the guanidinium group of arginine, but the presence of the methyl group would probably diminish the affinity of the AT-hook with DNA. As shown in Fig. 5, the methyl group will turn out to be in an external position if the hydrogen bonds between DNA and the guanidinium group are maintained. Contact of the hydrophobic methyl group with the solvent is in principle a destabilizing effect.

References

- Reeves R, Nissen MS (1990) The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J Biol Chem* 265: 8573–8582.
- Reeves R (2001) Molecular biology of HMGA proteins: hubs of nuclear function. *Gene* 277: 63–81.
- Sgarra R, Zammiti S, Lo Sardo A, Maurizio E, Arnoldo L, et al. (2010) HMGA molecular network: From transcriptional regulation to chromatin remodeling. *Biochim Biophys Acta* 1799: 37–47.
- Cleyen I, Van de Ven WJ (2008) The HMGA proteins: a myriad of functions (Review). *Int J Oncol* 32: 289–305.
- Foti D, Chiefari E, Fedele M, Iuliano R, Brunetti L, et al. (2005) Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice. *Nat Med* 11: 765–773.
- Hock R, Furusawa T, Ueda T, Bustin M (2007) HMG chromosomal proteins in development and disease. *Trends Cell Biol* 17: 72–79.
- Fusco A, Fedele M (2007) Roles of HMGA proteins in cancer. *Nat Rev Cancer* 7: 899–910.
- Reeves R, Beckerbauer LM (2003) HMGA proteins as therapeutic drug targets. *Prog Cell Cycle Res* 5: 279–286.
- Geierstanger BH, Volkman BF, Kremer W, Wemmer DE (1994) Short peptide fragments derived from HMG-I/Y proteins bind specifically to the minor groove of DNA. *Biochemistry* 33: 5347–5355.
- Huth JR, Bewley CA, Nissen MS, Evans JN, Reeves R, et al. (1997) The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat Struct Biol* 4: 657–665.
- Dragan AI, Liggins JR, Crane-Robinson C, Privalov PL (2003) The energetics of specific binding of AT-hooks from HMGA1 to target DNA. *J Mol Biol* 327: 393–411.
- Chen B, Young J, Leng F (2010) DNA Bending by the Mammalian High-Mobility Group Protein AT Hook 2. *Biochemistry* 49: 1590–1595.
- Drew HR, Dickerson RE (1981) Structure of a B-DNA dodecamer. III. Geometry of hydration. *J Mol Biol* 151: 535–556.

Materials and Methods

The oligonucleotides and the peptide used in this study were prepared by conventional chemical synthesis. Crystals were obtained by vapor diffusion at 4°C using 5 mM MgCl₂, 25 mM ammonium acetate, 25 mM Tris-HCl pH 7.5, 5% MPD (2-methyl-2,4-pentanediol) and 0.25 mM NiCl₂. The concentrations of DNA and peptide were 0.2 mM and 0.4 mM respectively, with a 1:2 ratio of DNA duplex to peptide. Crystals were obtained by increasing the concentration of MPD up to 48%. Diffraction data were collected at 110 K on beam line BM16 at the ESRF and processed with HKL2000 [21]. The structure was solved by molecular replacement with the program AMoRe [22], refined with REFMAC [23] and validated with 3DNA [24]. A bent DNA dodecamer (NDB code: BDL001) was used as a model in molecular replacement. Some of the amino acid residues showed very poor electron density. In particular Arg 33 could not be located in two of the complexes. Final refinement statistics and crystal data are given in Table 1. Attempts to obtain suitable crystals with many different oligonucleotides and AT-hook regions of HMGA1a have not been successful so far.

The crystals we have obtained have a unique unit cell, with four crystallographically independent complexes. The geometry of the unit cell is strongly related to the typical P2₁2₁2₁ structure of DNA dodecamers, described for the first time by Dickerson and co-workers [25]. The sides of our P2₁ unit cell coincide with the diagonals of the P2₁2₁2₁ crystals, so that our cell is equivalent to two unit cells of the Dickerson structure. A scheme of the asymmetric unit is given in Figure 6.

Figures have been prepared with either Pymol (available at <http://www.pymol.org/>) or with Cerius 2 (Accelrys Inc. San Diego, CA, USA).

Accession code

Protein data bank: Coordinates have been deposited with accession code 3UXW.

Supporting Information

Video S1 The video presents details of the structure of the AT-hook/DNA complex viewed from different angles and at different magnifications.

(MP4)

Author Contributions

Conceived and designed the experiments: JAS JLC NS. Performed the experiments: EF-S FA-R NS JP JAS JLC. Analyzed the data: EF-S FA-R NS JP JAS JLC. Contributed reagents/materials/analysis tools: EF-S FA-R NS JP JAS JLC. Wrote the paper: EF-S FA-R NS JP JAS JLC.

14. Soler-Lopez M, Malinina L, Liu J, Huynh-Dinh T, Subirana JA (1999) Water and ions in a high resolution structure of B-DNA. *J Biol Chem* 274: 23683–23686.
15. Dickerson RE, Goodsell DS, Kopka ML, Pjura PE (1987) The effect of crystal packing on oligonucleotide double helix structure. *J Biomol Struct Dyn* 5: 557–579.
16. Cui T, Leng F (2007) Specific recognition of AT-rich DNA sequences by the mammalian high mobility group protein AT-hook 2: a SELEX study. *Biochemistry* 46: 13059–13066.
17. Manabe T, Katayama T, Tohyama M (2009) HMGA1a recognition candidate DNA sequences in humans. *PLoS One* 4: e8004.
18. Bewley CA, Gronenborn AM, Clore GM (1998) Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annu Rev Biophys Biomol Struct* 27: 105–131.
19. Zhang Q, Wang Y (2010) HMG modifications and nuclear function. *Biochim Biophys Acta* 1799: 28–36.
20. Sgarra R, Diana F, Bellarosa C, Deklva V, Rustighi A, et al. (2003) During apoptosis of tumor cells HMGA1a protein undergoes methylation: Identification of the modification site by mass spectrometry. *Biochemistry* 42: 3575–3585.
21. Otwinowsky Z, Minor W (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol* 276: 307–326.
22. Navaza J (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallographica Section A* 50: 157–163.
23. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240–255.
24. Lu XJ, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31: 5108–5121.
25. Fratini AV, Kopka ML, Drew HR, Dickerson RE (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J Biol Chem* 257: 14686–14707.