Communication

Water and Ions in a High Resolution Structure of B-DNA*

(Received for publication, May 13, 1999, and in revised form, June 17, 1999)

Montserrat Soler-López‡, Lucy Malinina‡§, Jie Liu†, Tam Huynh-Dinh‡, and Juan A. Subirana‡

From the ‡Departament d’Enginyeria Química, Universitat Politècnica de Catalunya, Avinguda Diagonal 647, E-08028 Barcelona, Spain, the §Engelhardt Institute of Molecular Biology Russian Academy of Sciences, Vavilova 32, 117984 Moscow, Russia, and the ¶Unité de Chimie Organique, Institut Pasteur, Rue Dr. Roux 28, 75724 Paris, France

A detailed picture of hydration and counterion location in the B-DNA duplex d(GCGAATTCG) is presented. Detailed data have been obtained by single crystal x-ray diffraction at atomic resolution (0.89 Å) in the presence of Mg2+. The latter is the highest resolution ever obtained for a B-DNA oligonucleotide. Minor groove hydration is compared with that found in the Na+ and Ca2+ crystal forms of the related dodecamer d(CGCGAATTCCG). High resolution data (1.45 Å) of the Ca2+ form obtained in our laboratory are used for that purpose. The central GAATTC has a very stable hydration spine identical in all cases, independent of duplex length and crystallization conditions (centirotons, space group). However, the organization of the water molecules (tertiary and quaternary layers) associated with the central spine vary in each case.

When the first dodecamer structure in the B-form was established, a “spine of hydration” was found in the minor groove (1). Recent high resolution data on the same dodecamer (2–5) have shown that a series of water hexagons may build up on the spine of hydration. However, some of the water-water distances are too long to be considered hydrogen bonds, and they correspond rather to Van der Waals contacts. An additional result of the studies mentioned above is that either K+ (3) or Rb+ (5) may partially occupy the water sites. It has also been suggested (2) that Na+ may also occupy the water sites, but this suggestion has been challenged (4, 6).

The use of flash cooling and synchrotron radiation allows a much higher resolution in the x-ray diffraction data of oligonucleotides than was possible a few years ago. As a result, a much greater detail on water and ion distribution can be obtained.

The detailed study of water and ions around DNA is interesting in itself, but it is also relevant to understand protein-DNA interactions. It has been suggested (7, 8) that water-mediated polar contacts may contribute to the specificity of protein-DNA recognition. A better understanding of water structure around DNA is essential to ascertain the eventual role of ions and hydration waters in DNA interactions. The data obtained from high resolution oligonucleotide structures will be of great value in this sense.

Here we report the arrangement of water and divalent cations in the related structures of d(GCGAATTCG) in the presence of Mg2+ (resolution = 0.89 Å) and the calcium form of d(GCGAATTCG) (resolution = 1.45 Å). The high resolution achieved with these structures allows us to position with certainty many of the water and ions in the crystal. Thus, the influence of either monovalent (Na+) or divalent (Mg2+, Ca2+) cations can be ascertained. The structure of the water spine is very clear and allows us to determine which water molecules occupy fixed positions. Hydrogen bonds and Van der Waals contacts can be clearly distinguished.

The nonamer was crystallized, using a batch method, in sitting drops containing 0.5 mM DNA duplex, 1 mM acridine (Arg4) drug-peptide adduct, 20 mM sodium cacodylate buffer, pH 7, 100 mM MgCl2, and 35% MPD. Crystals grow in approximately 2 months to a typical size of 0.6 x 0.4 x 0.4 mm. The dodecamer crystallization is described elsewhere (9, 10). For data collection, crystals were mounted in a fiber loop and immediately flash-cooled at 120 K under a nitrogen vapor stream using an Oxford Cryosystems Cryostream. Data were collected using synchrotron radiation at EMBL beam line X11 in the DESY (Deutsches Elektronen-Synchrotron) Hamburg Outstation on a 345-mm MAR Research imaging-plate scanner. Three sets of data were collected at resolution cut-offs 0.89, 1.5, and 2.64 Å to avoid saturation of the high intensity reflections. The data were processed and reduced with DENOZ and SCALEPACK software packages (11).

The nonamer structure previously reported (12) at 2.05 Å resolution was used as a starting model. It was refined in XPLOR 3.8 (13) with a least-square target between 30 and 0.89 Å (Rfree = 0.248; Rfactor = 0.222). Simulated annealing protocols were employed. After running some cycles with XPLOR, we switched to SHEXL (14), where anisotropic refinement was carried out with some conjugate gradient cycles using the maximum resolution. The remaining structure was built into 2Fe–Fe electron density maps generated with SHELXPRO. The refinement converged at Rfactor = 0.145 for all data between 30 and 0.89 Å (R > 4σ). A very clear Cl− ion was found between two Mg2+ ions. Two poorly ordered sodium ions (B-factor = 241) near phosphate groups were tentatively assigned on the basis of several short distances (<2.6 Å) to neighbor water molecules. They were located on the major groove in regions where the solvent is not well ordered. They are not shown in the figures. The drug-peptide complex used in the crystallization was not found in the crystal. We expect to decrease the Rfactor by modeling the disordered regions.2 To our surprise only 110 water molecules could be precisely located, compared with 86 that were found in the structure reported

1 The abbreviations used are: MPD, 2-methyl-2,4-pentanediol; P, primary water layer; S, secondary water layer.
2 M. Soler-López, L. Malinina, and J. A. Subirana, work in progress.
Table I

Crystallographic and refinement statistics

|                  | Dodecamer | Nonamer |
|------------------|-----------|---------|
| Space group      | R3        | P 2_1 2_1 2_1 |
| a (Å)            | 41.12     | 21.95   |
| b (Å)            | 41.12     | 36.52   |
| c (Å)            | 99.91     | 52.73   |
| α = β (°)        | 90.00     | 90.00   |
| γ (°)            | 120.00    | 90.00   |
| Duplexes/asymmetric unit | 1       | 1       |
| Completeness     | 99.2%     | 97.6%   |
| Resolution range (Å) | 8–1.45   | 30–0.89 |
| Unique reflections | 11,005    | 32,684  |
| R_{sym} (overall/overall shell) | 0.053/0.337 | 0.059/0.300 |
| R_{free} (R > 4 σ) | 0.210 | 0.145 |
| R_{free} (all reflections) | 0.287 | 0.148 |
| No. DNA atoms    | 454       | 583     |
| No. ions         | 3.33 Ca^{2+}, 5 Mg^{2+}, 2 Na^{+}, 1 Cl^{-} | |
| No. water molecules | 144     | 144     |
| No. alternative occupancy waters | None | 5 |
| No. partial occupancy waters | None | 29 |
| Average B-factor (Å²) | 31.3/40.8 | 9.2/23.1 |

a R_{sym} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad b R_{free} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad c R_{sym} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad d R_{free} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad e R_{hklj} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad f R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad g R_{4} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad h R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad i R_{4} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad j R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad k R_{4} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad l R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad m R_{4} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad n R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad o R_{4} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad p R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad q R_{4} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad r R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad s R_{4} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad t R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad u R_{4} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad v R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad w R_{4} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad x R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad y R_{4} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad z R_{s} = \sum_{ij} |F_{ij}|-|F_{ij}| \text{,} \quad \text{Rsym was calculated using a random set containing 10% of observations that were omitted during refinement.} \text{,} \quad \text{Hydrogens are not included in the dodecamer.} \text{,} \quad \text{Divalent cations are given as large spheres, and water molecules are given as small spheres.} \text{,} \quad \text{Only the water molecules present inside the minor groove and those associated with ions are shown. Cl^{-} is represented as a larger sphere in the upper left corner. It is placed between two Mg^{2+} ions and interacts with the N4 atom of C2 and with waters associated with Mg^{2+} ions.} \text{,} \quad \text{The atomic resolution achieved with the nonamer data is (12) at 2.05 Å. An additional 5 molecules had closely spaced (<0.5 Å) alternative positions. Many other sites with partial occupancy appear to be present. No molecule of the MPD used for crystallization could be located. It appears that most of the rather large spaces available for solvent molecules are in a poorly ordered glassy state. However, there are regions that have a highly ordered water network, probably influenced by the presence of Mg^{2+} ions.} \text{,} \quad \text{Refinement details for the high resolution structure of the dodecamer are reported elsewhere (10). Crystal data and refinement statistics for both structures are listed in Table I. Stereo views are shown in Fig. 1. The hydration spine, counte- rions, and their associated waters are also shown in the figure. In the case of the dodecamer in the R3 space group, a detailed comparison with the standard P2_1 2_1 2_1 structure (2) is presented elsewhere (10).} \text{,} \quad \text{The atomic resolution achieved with the nonamer data is clearly evident in Fig. 2. The characteristic triplet interaction of this nonamer (12) is shown in the figure. Comparison of the duplex structure of d(GCGAATTCGCG) with the data obtained at a lower resolution (12) does not show any significant difference (root mean square excluding phosphates between both structures is 0.33 Å). The main difference is that in our study four phosphate groups are disordered and have at least a double conformation. In Fig. 1, only the major conformation is represented. Alternative phosphate positions are available at the reported Nucleic Acid Database file. The strong similarity between the two structures determined at different resolutions gives confidence in the data presently available on oligonucleotide structures that have been obtained mostly at resolutions around 2.5 Å (Ref. 2). This is itself an important conclusion of the work reported here, although the most interesting new data obtained relate to the organization of water and ions around DNA, as we will show below.} \text{,} \quad \text{In Fig. 3 we present the minor groove hydration for the Na^{+}, Ca^{2+}, and Mg^{2+}} \text{,} \quad \text{dodecamer. All three spines have been superimposed by using waters P2, P3, and P4 in the primary hydration layer and S3 and S4 in the secondary layer. The nomenclature of Shui et al. (3) is used. They distinguish four hydration layers at increasing distances from the bottom of the minor groove. From Fig. 3 it is obvious that the superposition of the five water molecules mentioned above is excellent. Waters at either end of the spine (P1, P5) are no longer so well superimposed. Oligonucleotides of different lengths crystallized with different cations (Na^{+}, Mg^{2+}, Ca^{2+}) in different space groups (P2_1 2_1 2_1, R3) are compared in the figure. It is evident that the central AATT sequence contains a quite rigid spine of hydration, first described by Drew and Dickerson (1), which is very stable. The minor groove is also very narrow in all three cases, although in the nonamer case it is about 0.6 Å wider at the center, as already reported (12).} \text{,} \quad \text{A tertiary layer of hydration is also clear in all cases. Each water molecule in the secondary layer has an additional hydrogen-bonded water molecule. However, the orientation of water molecules in the tertiary layer with respect to those in the secondary layer is different in all three cases. For example, in}
the nonamer case the S2 water has two associated waters instead of one in the tertiary layer.

The quaternary layer of hydration is quite different in all cases. Many of the waters are not hydrogen bonded with those of the tertiary layer as evident in Fig. 3. In fact, they are associated with chains of waters that interact with the phosphate groups of the same and neighbor duplexes. They should not be considered an integral part of the minor groove water spine because their position depends on the interactions in the crystal. In the nonamer crystal structure, the position of water molecules near the P1 side of the spine is very well defined, whereas they are mainly disordered on the P4/P5 side. Such differences are due to the fact that the two ends of the duplexes lie in different regions of the crystal and have different interactions with neighbor duplexes. The water molecules associated with the P1 end of the minor groove water spine form a system of polygons and solvent chains that interconnect three nonamer molecules and several ions (Mg$^{2+}$, Cl$^{-}$) in the crystal.

The ends of the water spine (sites P1 and P5) are also very different in the three cases shown in Fig. 3, due in part to the presence of divalent cations in these regions. In the case of the nonamer, for example, instead of a single P1 water molecule there are two water molecules at this level, one interacting

**FIG. 2.** Electron density map of one of the triplets in the nonamer. The triplet is formed between G10 and the G2-C18 base pair of a neighboring molecule. The values of $2F_o - F_c$ contoured at 1.5$\sigma$ are shown. The atomic resolution of the model can be clearly appreciated.

**FIG. 3.** Comparison of the water spine in three different structures. The water spines and associated waters lie on a slightly twisted plane. Waters in each spine are given in different colors; they are shown in a–c projected onto the average plane. a, d(GCGGAATTCGCG) crystallized in space group P2,2,2, in a sodium buffer (from Ref. 2); b, the same dodecamer crystallized in space group R3 in a calcium buffer; c, d(GCGGAATTCG) in a magnesium buffer; d, superposition of the three water spines associated with the central GAATTC sequence (coordinates given in Ref. 2 are used for the duplex). The three spines have been superimposed at the P2-S3-P3-S4-P4 central waters, which occupy very similar positions. The latter five water molecules are represented in orange in the three cases. Continuous bonds indicate distances shorter than 3.4 Å (hydrogen bonds). The nomenclature of the water molecules has been taken from Ref. 3. P indicates the primary water layer and S the secondary layer. For clarity, waters in the quaternary layer are omitted in d. Some of them only have van der Waals contacts with the other water molecules (isolated spheres in a and c). Divalent cations are shown in blue (Mg$^{2+}$ light blue; Ca$^{2+}$, dark blue). Hydrogen bonds with the duplex are not shown.
with cytosine and the other with guanine.

The results presented demonstrate that the GAATTTC sequence has a very well defined water spine associated with the narrow minor groove of the AATT central region. This water spine has two very well defined and constant layers, with a somewhat variable third layer. On top of the latter layer, a system of water molecules defines a set of pentagons/hexagons and chains of water molecules. The exact geometry depends on the sequence, the length of the duplex, and the crystallization conditions used in each case. The presence of strongly associated water molecules in the A/A/T/T and AT base steps may contribute to stabilizing the conformational features of these base steps, which are known to vary very little (15, 16) in different oligonucleotides. Thus, to interpret the detailed structural features of different DNA sequences, the roles of water and ions should be taken into account.

Finally, we should discuss the role of the divalent cations in the structure of the duplexes. The high resolution structure of d(GCGAATTCG) includes several ions (Table I). Because the d(GCGAATTCG) duplex has 16 negative phosphate charges, some charges remain to be neutralized in each asymmetric unit, probably by additional disordered Mg$^{2+}$/Na$^+$ ions that have not been detected. In fact, some electron density peaks that have been assigned to water molecules may correspond to such ions because they show some short contacts. The Mg$^{2+}$ ions are mainly involved in the interaction among different duplexes in the crystal, but none of them are found in the region of the minor groove water spine. Instead, two Mg$^{2+}$ ions are found at both ends of the minor groove and in fact interact with the ends of the water spine, as shown in Fig. 3. A similar situation is found in the R3 structure, where Ca$^{2+}$ ions are present at similar positions, as shown in Figs. 1 and 3.

Although the minor groove is considered to be (17) a very electronegative region, it appears that the geometry of the AATT region with its associated water spine allows neither hydrated Ca$^{2+}$ nor Mg$^{2+}$ to penetrate inside the minor groove. However, in oligonucleotides with a different sequence (18–20), which have a wider minor groove, hydrated divalent cations do penetrate into the minor groove. The influence of sequence has also been confirmed by NMR experiments showing that hydrated Mn$^{2+}$ ions do not penetrate the A$_1$T$_4$ minor groove but do penetrate into T$_4$A$_1$ (21).

The eventual substitution of some of the molecules in the water spine by Na$^+$ ions has been a question of debate (2, 4–6, 22), although it appears that either Rb$^+$ (5) or K$^+$ (3) may indeed replace, in part, some of the water molecules. In fact, the latter ions have a larger ionic radius more suitable than Na$^+$ for the geometry of the minor groove.

In summary, high resolution studies allow an improved knowledge of ion and water distribution around oligonucleotides. However, the relationship between water structure after flash-cooling and water structure at room temperature is not clear, and further studies are required in this direction. In fact, we found that only a limited number of solvent molecules occupies well defined sites, and most of the waters appear to be in a disordered, glassy state. Some of the counterions could not be located, because the oligonucleotide negative charges are not fully neutralized. Also no MPD molecule was found to occupy any definite position. In fact no MPD molecule has been found in any oligonucleotide crystal. In the protein crambin, also studied (23) at high resolution (0.83 Å); and only one ethanol molecule could be located, although crystals were obtained around 70% ethanol concentration. The absence of MPD in this and in all the reported oligonucleotide structures may indicate that either MPD is excluded from the crystals or is always disordered.

Acknowledgments—We thank Drs. G. DeTitta and J. Luft for advice on crystallization and Drs. V. Tereshko and L. Williams for sending manuscripts before publication.

REFERENCES