Biochemical and Biophysical Research Communications **283,** 219–223 (2001)
doi:10.1006/bbrc.2001.4755, available online at http://www.idealibrary.com on IDEAL®

Tertiary Structures of the *Escherichia coli* and Human Chromosome 21 Molecules of DNA

Petr Hanzálek and Jaroslav Kypr[1]
Institute of Biophysics of the Academy of Sciences of the Czech Republic, Královopolská 135, CZ-61265 Brno, Czech Republic

Received March 27, 2001

Using the NDB database, we calculated geometrical parameters that were needed to reproduce crystal structures of short DNA fragments in a phosphorus atom representation. The geometrical parameters were included in a software generating tertiary structures of, for example, the *Escherichia coli* and human chromosome 21 molecules of DNA whose complete nucleotide sequences are deposited in the EMBL and related databases. Both molecules were found to be heavily folded and composed of domains. A more elaborate version of the present approach will make analysis and comparison possible of tertiary structures of genomic DNA molecules of various chromosomes to identify the chromosome evolutionary and functional relationships.

Key Words: **DNA crystal structures; phosphorus atom representation; genomic DNA molecules; tertiary structures.**

Life crucially depends on molecular operations that take place in the cell nucleus where two meters of human DNA are heavily compacted (reviewed e.g., in Ref. 1). The compaction is facilitated by chromosomal proteins but known properties of DNA make it unlikely that DNA is a passive partner in the compaction process (2). Rather we would say that free chromosomal molecules of DNA adopt a compact tertiary structure which, to a certain extent, determines the three-dimensional structure of the whole chromosome. However, determination of the tertiary structures of free chromosomal molecules of DNA is very difficult because the megabase and still longer molecules of chromosomal DNA are sheared even due to DNA solution shaking. Another problem is the molecule visualization because deposition of DNA on a surface, which is inevitable in most kinds of microscopy, potentially should modify, if not destroy, the native tertiary structure of DNA. Here we overcome these obstacles by a computer approach described below.

Protein tertiary structures are greatly simplified by the alpha carbon representation that is very useful and instructive on many occasions (3–5). A logical analogy is the phosphorus atom representation in nucleic acids (6). We have calculated phosphorus atom distances, and the pseudovalence and pseudodihedral angles from the NDB database (7) of the B-DNA and A-DNA crystal structures, and reproduced the crystal structures of DNA using these parameters. This work is extensive and will be described in detail elsewhere (P. Hanzálek and J. Kypr, in preparation). Here we extend this work to megabase molecules of genomic or chromosomal DNA whose complete nucleotide sequences are deposited in the EMBL (8) and related databases. Our approach makes it possible that phosphorus atom representations are generated not only of the tertiary structures of the whole bacterial genomes but also of the human chromosome molecules of DNA having dozens of megabases in length. The first results obtained with these megabase molecules of DNA are reported below.

MATERIALS AND METHODS

The complete nucleotide sequence of the *E. coli* DNA was determined by Blattner *et al.* (9). The (almost) complete nucleotide sequence of the human chromosome 21 DNA was determined by Hattori *et al.* (10). The Cartesian coordinates of the phosphorus atoms are calculated by our program GENOMesHAPE written in C++ that uses the nucleotide sequence of the DNA molecule as the input data. The program is based on the known algorithm for a generation of a spatial curve using distances of the neighboring phosphorus atoms, the pseudovalence angles of three consecutive phosphorus atoms and the pseudodihedral angles of the four consecutive phosphorus atoms along the polynucleotide chain (11, 12). The geometrical parameters were separately determined for A-DNA and B-DNA so that we can generate B-form as well as A-form of the genomic DNA molecules. The calculated Cartesian coordinates of the phosphorus atoms then go to MOLMOL (13) that produces the molecule pictures shown below. The calculations were carried out with IBM-PC (two Pentium II processors 400 MHz, 512 MB memory) operating under Red Hat Linux 6.1.

[1] To whom correspondence should be addressed. Fax: +4205 4124 0497. E-mail: kypr@ibp.cz.

0006-291X/01 $35.00

TABLE 1

Intrastrand Distances of the Neighboring Phosphorus Atoms P Which Belong to the Indicated Nucleotides in the Crystal Structures of the A-Form and B-Form DNA

Nucleotide	A-form (Å)	B-form (Å)
pAp	6.33	6.73
pCp	5.95	6.47
pGp	6.14	6.86
pTp	6.12	6.73
pNp	6.05	6.72

RESULTS

Tables 1–3 respectively summarize the P–P distances, P–P–P pseudovalence angles and the P–P–P–P pseudodihedral angles used for the present modeling. These values were assigned to the nucleotide sequence of the modeled DNA. If the tables contained no value for a dinucleotide or trinucleotide, then we used the values for twofold degenerate dinucleotides and trinucleotides, where R = A or G and Y = C or T. If even these values were missing in the tables, then we used the average values (the bottom row in each table). The nucleotide sequence of the *E. coli* genome is complete (9) but that of the human chromosome 21 is not (10) so that we assigned the average values to the unspecified nucleotides in the sequence as well. Although Tables 2 and 3 are still incomplete because the crystal structures of mainly B-DNA do not contain all possible dinucleotides and especially trinucleotides, the data are nonetheless sufficient to reproduce the crystal structures of DNA quite well in the phosphorus atom representation (Fig. 1 and Hanzálek and Kypr, in preparation). This motivated us to use the data for modeling of megabase molecules of genomic DNA.

TABLE 2

Intrastrand Pseudovalence Angles (in Radians) of Three Consecutive Phosphorus Atoms P Which Belong to the Indicated Dinucleotides in the Crystal Structures of A-Form and B-Form DNA

Dinucleotide	A-form	B-form
pApAp	—	2.53
pApCp	2.62	—
pCpAp	2.76	—
pCpCp	2.63	—
pCpGp	2.58	—
pGpCp	2.63	—
pGpGp	2.62	—
pGpTp	2.73	—
pTpGp	2.60	—
pTpTp	—	2.56
pRpRp	2.62	2.53
pRpYp	2.64	—
pYpRp	2.61	—
pYpYp	2.64	2.56
pNpNp	2.63	2.55

TABLE 3

Intrastrand Pseudodihedral Angles (in Radians) of the Four Consecutive Phosphorus Atoms P in the Crystal Structures of A-Form and B-Form DNA

Trinucleotide	A-form	B-form
pApApAp	—	1.19
pCpCpCp	1.21	—
pGpGpGp	1.21	—
pTpTpTp	—	1.03
pGpCpGp	1.25	—
pGpTpGp	1.38	—
pCpGpCp	1.33	—
pCpApCp	1.21	—
pGpGpCp	1.34	—
pGpGpTp	1.29	—
pCpCpGp	1.33	—
pCpCpAp	0.77	—
pApCpCp	0.63	—
pGpCpCp	1.33	—
pTpGpGp	1.43	—
pCpGpGp	1.42	—
pRpRpRp	1.21	—
pRpRpYp	1.19	—
pRpYpRp	1.27	—
pRpYpYp	1.13	—
pYpRpRp	1.43	—
pYpRpYp	1.32	—
pYpYpRp	1.16	—
pYpYpYp	1.21	—
pNpNpNp	1.24	1.11

Figure 2 presents tertiary structure of the whole DNA molecule of *Escherichia coli* generated as outlined in the previous paragraph and under Materials and Methods. Every 500th phosphorus atom is only shown for clarity, i.e., each dot represents about 50 double helix turns. The native *E. coli* DNA is circular, but we cannot generate DNA circles with the current version of our software. Yet Fig. 1 shows that the molecule ends are fairly close to each other so that the linear molecule shape will not change much upon circularization. The molecule tertiary structure is remarkably compact. The DNA length is about 1.5 mm but it only occupies a sphere having 0.034 mm in diameter. Hence the compaction ratio (length divided by the sphere radius) is about 50 in the absence of any protein. The *Escherichia coli* molecule of DNA exhibits a domain structure. It seems to be composed of three domains and each domain is furthermore composed of many highly compact subdomains. Some of the subdomains are loose and extrude from the bulk of DNA to be readily accessible. These properties are typical of the bacterial nucleoid (14) including the *E. coli* DNA (15) as well as the eukaryotic chromosomes that are also composed of domains and loops (16–18). Circles or loops were also predicted to occur in the HIV provirus DNA (19).

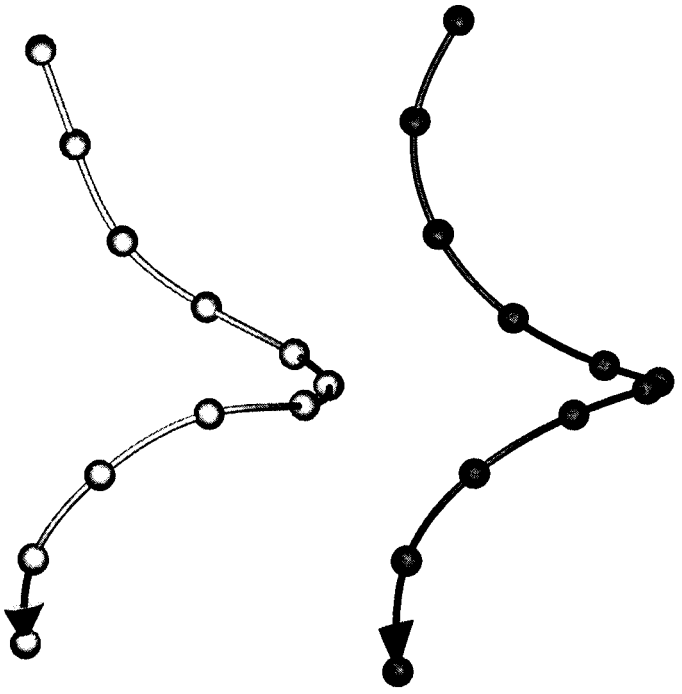

FIG. 1. A representative example of how the present approach reproduces the crystal structure of a strand of d(GCGCTTAAGCGC) (B-DNA, NDB code bdl002) in the phosphorus atom representation. The crystal structure is left, the model is right. The small spheres represent the phosphorus atoms.

Figure 2 was generated using the information extracted from the crystal structures of B-DNA. However we have calculated an analogous information for A-DNA as well so that our software can also generate A-form of the *E. coli* DNA. This A-form is shown in Fig. 3. It differs substantially from the B-form not only by the global shape and the domain organization, but also by the higher degree of the local compactness. This is a general property, i.e., the A-forms of genomic DNA molecules are generally more compact than the B-forms (Hanzálek and Kypr, in preparation). The B-form is believed to be the predominant conformation but DNA is dehydrated *in vivo* (20) and the dehydration stabilizes the A-form (21). For example, DNA is an A-form in dormant spores of *Bacillus* species (22). Hence we generate and analyze both these basic conformations of DNA.

The DNA of human chromosome 21 is more than 6 times longer than the DNA of *Escherichia coli*. Yet we can still generate its three dimensional shape though this is already close to the maximum DNA length we can manage using our current hardware. The B-form of the DNA molecule of the human chromosome 21 is depicted in Fig. 4. Each 3000th phosphorus atom is only shown to make the molecule synoptical. Figure 4 demonstrates that majority of DNA is condensed in one domain whereas the remainder generates a few separate small domains. It is interesting that the beginning of the molecule (the first approx 0.75 megabase, containing the telomere) is well separated from the subtelomeric region by a more or less straight long connector. The subtelomeric region is then separated from the major domain by another long piece of noncompacted DNA. These properties will be subject to a more detailed analysis in a forthcoming work.

A-form of the human chromosome 21 molecule of DNA is very different from the B-form of the same molecule again (not shown). The central large domain is unfolded and the molecule generates a number of smaller, well separated domains. The human chromosome 21 DNA is about 1cm in length and it occupies a sphere having 58 and 84 micrometers in diameter in the A-form and B-form, respectively. Hence this DNA is compacted more than 10,000 times. Our approach makes possible to look at any molecule with a desired resolution, i.e., we can visualize e.g., every 10,000th, 1000th, 100th, or every 10th phosphorus atom (of course, the latter two, i.e., the high "resolutions" will require to "split" the molecule into fragments of appropriate length to make it manageable with our soft-

FIG. 2. Tertiary structure of linearized *E. coli* DNA. The symbol (●) indicates ends of the molecule. This structure was generated using the intrastrand distances of the successive phosphorus atoms and their pseudovalence and pseudodihedral angles occurring in the crystal structures of B-DNA. Every 500th phosphorus atom is represented by a dot. This figure was prepared with the program MOLMOL (Koradi *et al.*, 1996). For details, see the main text and Materials and Methods.

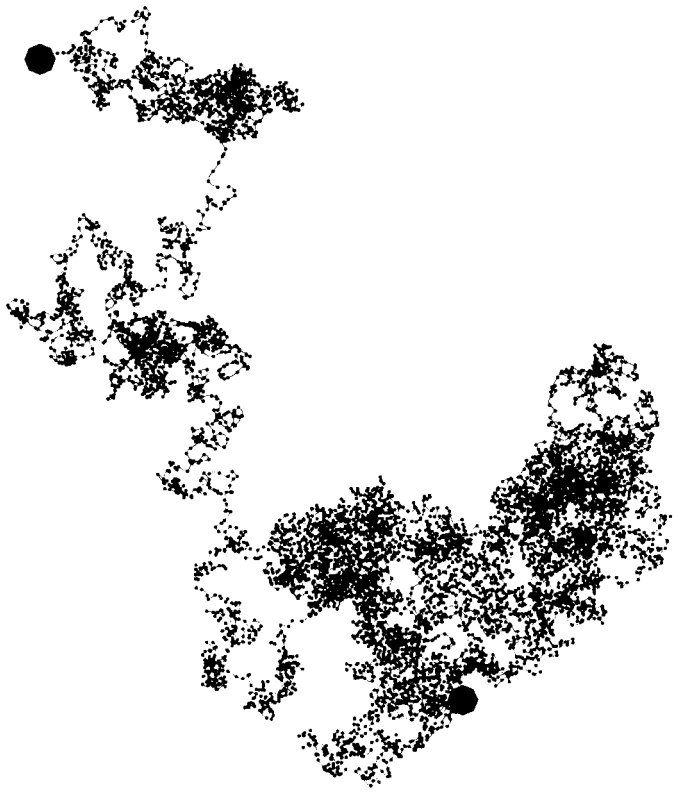

FIG. 3. A-form of linearized *E. coli* DNA. The symbol (●) indicates ends of the molecule. This figure was prepared with the program MOLMOL (Koradi *et al.,* 1996).

ware). The possibility to choose any subset of the phosphorus atoms for the visualization and further analysis, will make it possible to study hierarchical organization of the tertiary structures of DNA molecules of each human chromosome.

DISCUSSION

Tertiary structures are well understood with globular proteins because hundreds of them have already been successfully crystallized and analyzed by X-ray diffraction. In contrast, the tertiary structures of native molecules of DNA are not understood at all because of their huge length and fibrous nature. That is why we have designed the present approach to study the tertiary structures of megabase molecules of genomic or chromosomal DNA. In the current simplest version, our method relies on the average geometrical information derived from the available crystal structures of DNA. This avoids the notorious pitfalls of current methods of DNA modeling because all forces, the entropic effects, solvent and ions are adequately included in our approach. Our simulations only extrapolate experimental data, i.e., the DNA crystal structures. They are sufficiently simple to be used for tertiary structure modeling of DNA molecules that have even hundreds of megabases in length.

The only artifacts the present approach can produce, would originate from the pitfalls that the crystal structures of DNA suffer from. In this direction, dehydration is the most important because DNA is generally dehydrated in the crystalline state (23, 24). At first sight, this seems to be a bad news, but in fact it is not because DNA is heavily dehydrated in the cell nucleus as well (20) so that the crystal structures probably reflect the situation *in vivo* better than the dilute aqueous solutions. In the current version of the software, we use the average values of the geometrical parameters observed in the crystals, which means that all inaccuracies and other effects, having a random nature, are eliminated.

FIG. 4. B-form of the DNA molecule of human chromosome 21. The symbol (●) indicates ends of the molecule. This figure was prepared with the program MOLMOL (Koradi *et al.,* 1996).

In conclusion, we would like to point out that the structures presented here are only the first approximation of what our method can provide. Yet even the first approximation suggests properties that appear to be realistic and biologically relevant. However, the major message of this communication is the demonstration that the current NDB and EMBL databases already contain data to model tertiary structures of the whole genomic or chromosomal molecules of DNA.

ACKNOWLEDGMENT

This work was supported by Grant A5004802 awarded to Jaroslav Kypr by the Grant Agency of the Academy of Sciences of the Czech Republic.

REFERENCES

1. Cook, P. (1998) *Science* **281,** 1466–1467.
2. Roux-Rouquie, M., and Marilley, M. (2000) *Nucleic Acids Res.* **28,** 3433–3441.
3. Oldfield, T. J., and Hubbard, R. E. (1994) *Proteins Struct., Funct., and Genet.* **18,** 324–337.
4. Flocco, M. M., Mowbray, S. L. (1995) *Protein Sci.* **4,** 2118–2122.
5. Reese, M. G., Lund, O., Bohr, J., *et al.* (1996) *Protein Eng.* **9,** 733–740.
6. Tung, Ch.-S., and Soumpasis, D. M. (1996) *Biophys. J.* **70,** 917–923.
7. Berman, H. M., Gelbin, A., and Westbrook, J. (1996) *Prog. Biophys. Mol. Biol.* **66,** 255–288.
8. Stoesser, G., Baker, W., van den Broek, A., *et al.* (2001) *Nucleic Acids Res.* **29,** 17–21.
9. Blattner, F. R., Plunkett, G., III, Bloch, C. A., *et al.* (1997) *Science* **277,** 1453–1474.
10. Hattori, M., Fujiyama, A., Taylor, T. D., *et al.* (2000) *Nature* **405,** 311–319.
11. Colombano, S., Rein, R., and MacElroy, R. D. (1980) *Comp. Prog. Biomed.* **11,** 3–8.
12. Sugeta, H., and Miyazawa, T. (1967) *Biopolymers* **5,** 673–679.
13. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics* **14,** 51–55.
14. Robinow, C., and Kellenberger, E. (1994) *Microbiol. Rev.* **58,** 211–232.
15. Niki, H., Yamaichi, Y., and Hiraga, S. (2000) *Genes Dev.* **14,** 212–223.
16. Agard, D. A., and Sedat, J. W. (1983) *Nature* **302,** 676–681.
17. Bickmore, W. A., and Oghene, K. (1996) *Cell* **84,** 95–104.
18. Münkel, Ch., Eils, R., Dietzel, S., *et al.* (1998) *J. Mol. Biol.* **285,** 1053–1065.
19. Albert, F. G., Bronson, E. C., Fitzgerald, D. J., and Anderson, J. N. (1995) *J. Biol. Chem.* **270,** 23570–23581.
20. Hildebrandt, E. R., and Cozzarelli, N. R. (1995) *Cell* **81,** 331–340.
21. Franklin, R. E., and Gosling, R. G. (1953) *Acta Crystallogr.* **6,** 673–677.
22. Setlow, P. (1992) *Mol. Microbiol.* **6,** 563–567.
23. Vorlíčková, M., Subirana, J. A., Chládková, J., Tejralová, I., Huynh-Dinh, T., Arnold, L., and Kypr, J. (1996) *Biophys. J.* **71,** 1530–1538.
24. Kypr, J., Chládková, J., Zimulová, M., and Vorlíčková, M. (1999) *Nucleic Acids Res.* **27,** 3466–3473.