BIOINFORMATICS APPLICATIONS NOTE Vol. 16 no. 11 2000
Pages 1044–1045

model.it: building three dimensional DNA models from sequence data

Kristian Vlahoviček and Sándor Pongor

Protein Structure and Function Group, International Centre for Genetic Engineering and Biotechnology, Area Science Park, 34012 Trieste, Italy

Received on May 17, 2000; revised on July 11, 2000; accepted on July 19, 2000

Abstract

Summary: *A WWW server is described for creating 3D models of canonical or bent DNA starting from sequence data. Predicted DNA trajectory is first computed based on a choice of di- and tri-nucleotide models (M.G.Munteanu* et al.*,* Trends Biochem. Sci. ***23****, 341–347, 1998); an atomic model is then constructed and optionally energy-minimized with constrained molecular dynamics. The data are presented as a standard PDB file, directly viewable on the user's PC using any molecule manipulation program.*

Availability: *The* model.it *server is freely available at http://www.icgeb.trieste.it/dna/*

Contact: *kristian@icgeb.trieste.it; pongor@icgeb.trieste.it*

Supplementary information: *A series of help files is available at the above address.*

Visualization of DNA 3D structure is essential for designing and interpreting molecular biology experiments. Several PC programs are available for modelling canonical and curved DNA (Bansal *et al.*, 1995; Dlakic and Harrington, 1998; Jursa, 1994; Shpigelman *et al.*, 1993; Tung and Carter, 1994), but very often these programs are built for a specific computer platform and use custom user interfaces that may be too complex for an intermittent user. Moreover, PC programs have difficulties in building curved DNA models, as the construction of a sugar phosphate backbone requires compute-intensive molecular mechanics calculations. The *model.it* server was designed to provide 3D models of DNA in response to DNA sequence queries. The results are presented as a standard PDB file that can be viewed directly using any of the widely available molecule manipulation programs such as Swiss PDB viewer (Guex *et al.*, 1999) or Rasmol (Sayle and Milner-White, 1995). The server's help files contain instructions for installing these programs.

The *model.it* server was written using 'NAB'—a high level molecule manipulation language (Macke and Case, 1998). The construction is based on canonical A- and B-DNA basepairs with idealized geometries. The molecule building process is performed in a local coordinate system based on transformations defined by the Cambridge convention (Dickerson, 1989). The first basepair is placed at the origin of a Cartesian coordinate system and then each following basepair is rotated (roll, tilt and twist rotations) and translated (rise is 3.4 Å for B- and 2.8 Å for A-DNA, respectively) with respect to the previous one. The server provides a choice of dinucleotide (Bolshoy *et al.*, 1991; Olson *et al.*, 1993; Ulyanov and James, 1995) and trinucleotide (Brukner *et al.*, 1995) parameter sets for building curved DNA. Coordinates of the sugar–phosphate backbone are optionally optimized with constrained molecular dynamics using energy parameters from the AMBER package (Case *et al.*, 1997). At present, the server can produce models of 700 bp in length, but models longer than 50 bp will not be optimized. Modelling of canonical, straight B or A DNA structures proceeds in a similar way, but without the need for backbone geometry optimization.

Naturally, modelling of DNA based on idealized or statistically derived geometry parameters is a rough approximation that does not take into account interactions with proteins or with the medium. While these models may not be highly accurate in terms of atomic detail, the gross spatial orientation of motifs with respect to each other is still a useful piece of information for designing or explaining experimental data. Since the DNA segments in question are often too long for current methods of structure determination, structure prediction can be particularly useful in this sense. Moreover, location of bends in DNA sequences can be predicted with reasonable accuracy, thus the orientation of motifs with respect to adjacent curved elements can also be studied. For example, it is an intriguing question whether or not such features can remain conserved in functionally equivalent sequences that have little sequence homology. Figure 1 shows a non-redundant set of *Zea mays* promoter regions chosen from the Eukaryotic Promoter Database (Perier *et al.*, 1999) as an example. Their predicted conformations show that the corresponding DNA segments are all curved to the same extent and 11 out of 14 are bent in the same direction.

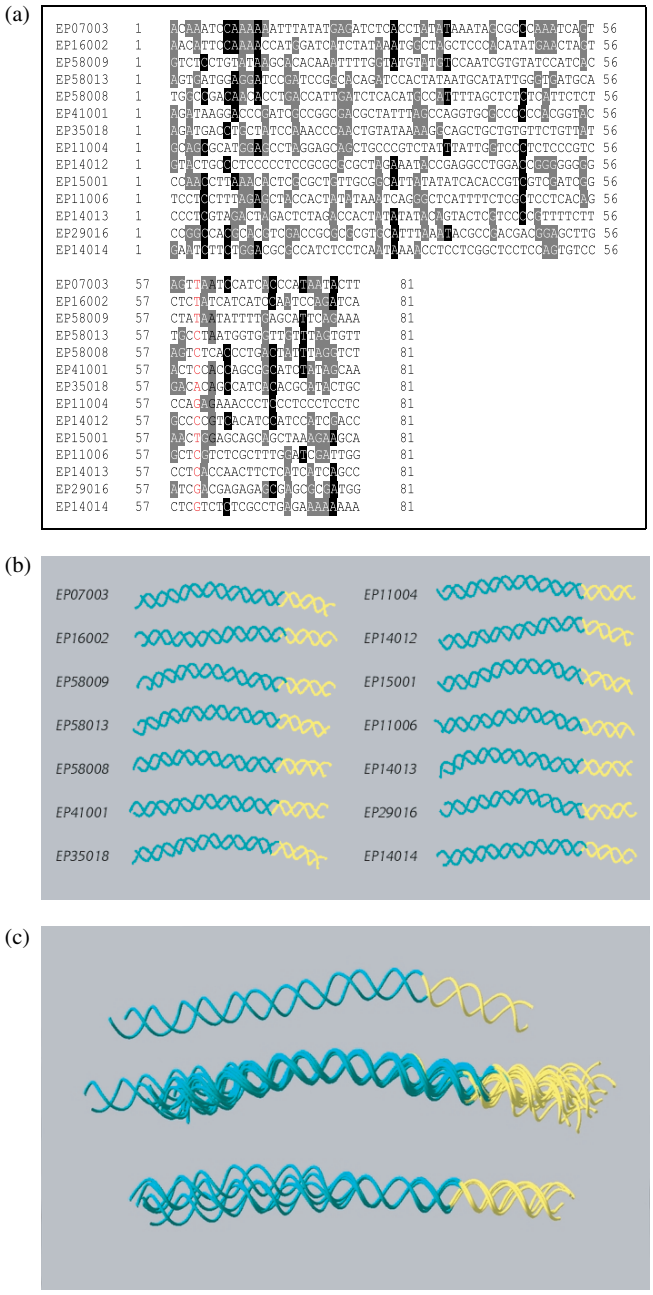

Fig. 1. (a) Alignment of 14 *Zea mays* promoter regions from EPD database. Transcription origin is marked with a red letter. (b) Predicted conformation based on the consensus parameter set (see Ref. Munteanu *et al.*, 1998, for details). Transcribed part is shown in yellow. (c) Superposition reveals three conformation groups.

References

Bansal,M., Bhattacharyya,D. and Ravi,B. (1995) Nuparm and Nucgen: software for analysis and generation of sequence dependent nucleic acid structures. *Comput. Appl. Biosci.*, **11**, 281–287.

Bolshoy,A., McNamara,P., Harrington,R.E. and Trifonov,E.N. (1991) Curved DNA without a-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. USA*, **88**, 2312–2316.

Brukner,I., Sanchez,R., Suck,D. and Pongor,S. (1995) Sequence-dependent bending propensity of DNA as revealed by Dnase I: parameters for trinucleotides. *Embo J.*, **14**, 1812–1818.

Case,D.A., Pearlman,D.A., Caldwell,J.W., III,T.E.C., Ross,W.S., Simmerling,C.L., Darden,T.A., Merz,K.M., Stanton,R.V., Cheng,A.L., Vincent,J.J., Crowley,M., Ferguson,D.M., Radmer,R.J., Seibel,G.L., Singh,U.C., Weiner,P.K. and Kollman,P.A. (1997) *Amber 5*. University of California, San Francisco.

Dickerson,R.E. (1989) Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.*, **17**, 1797–1803.

Dlakic,M. and Harrington,R.E. (1998) Diamod: display and modeling of DNA bending. *Bioinformatics*, **14**, 326–331.

Guex,N., Diemand,A. and Peitsch,M.C. (1999) Protein modelling for all. *Trends Biochem. Sci.*, **24**, 364–367.

Jursa,J. (1994) DNA modeller: an interactive program for modelling stacks of DNA base pairs on a microcomputer. *Comput. Appl. Biosci.*, **10**, 61–65.

Macke,T. and Case,D.A. (1998) Modeling unusual nucleic acid structures. In Leontes,N.B. and SantaLucia,J. (eds), *Molecular Modeling of Nucleic Acids* American Chemical Society, Washington DC, pp. 379–393.

Munteanu,M.G., Vlahovicek,K., Parthasarathy,S., Simon,I. and Pongor,S. (1998) Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. *Trends Biochem. Sci.*, **23**, 341–347.

Olson,W.K., Marky,N.L., Jernigan,R.L. and Zhurkin,V.B. (1993) Influence of fluctuations on DNA curvature. A comparison of flexible and static wedge models of intrinsically bent DNA. *J. Mol. Biol.*, **232**, 530–554.

Perier,R.C., Junier,T., Bonnard,C. and Bucher,P. (1999) The eukaryotic promoter database (Epd): recent developments. *Nucleic Acids Res.*, **27**, 307–309.

Sayle,R.A. and Milner-White,E.J. (1995) Rasmol: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374.

Shpigelman,E.S., Trifonov,E.N. and Bolshoy,A. (1993) Curvature: software for the analysis of curved DNA. *Comput. Appl. Biosci.*, **9**, 435–440.

Tung,C.S. and Carter,E.S. (1994) Nucleic acid modeling tool (Namot): an interactive graphic tool for modeling nucleic acid structures. *Comput. Appl. Biosci.*, **10**, 427–433.

Ulyanov,N.B. and James,T.L. (1995) Statistical analysis of DNA duplex structural features. *Meth. Enzymol.*, **261**, 90–120.