The High-Resolution Crystal Structure of a Parallel-Stranded Guanine Tetraplex

Gerard Laughlan; Alastair I. H. Murchie; David G. Norman; Madeleine H. Moore; Peter C. E. Moody; David M. J. Lilley; Ben Luisi

Science, New Series, Vol. 265, No. 5171 (Jul. 22, 1994), 520-524.

Stable URL:
http://links.jstor.org/sici?sici=0036-8075%2819940722%293%3A265%3A5171%3C520%3ATHCSOA%3E2.0.CO%3B2-8

Science is currently published by American Association for the Advancement of Science.

tomeres, consistent with our finding that it is the central, nonmarginal region of the blastula that undergoes extensive cell mixing during epiboly. Strehlow and Gilbert (*5*) observed that the descendants of each eight-cell blastomere gave rise to a limited array of mesodermal tissues. These tissues arise at the margin, where we have observed limited cell mixing.

In sum, the descendants of each cleavage blastomere reside in a limited region of the gastrula and, therefore, populate only a subset of the possible cell fates. The particular subset of fates expressed depends on the position of the original blastomere relative to the future dorsoventral axis of the embryo. Because the dorsoventral axis is not fixed with respect to the early cleavage planes and the scatter of clones varies from embryo to embryo, the specific fate of any blastomere cannot be predicted at the cleavage stage.

REFERENCES AND NOTES

1. C. B. Kimmel and R. M. Warga, *Science* **231**, 365 (1986).
2. _____ and T. F. Schilling, *Development* **108**, 581 (1990).
3. C. B. Kimmel and R. D. Law, *Dev. Biol.* **108**, 94 (1985).
4. R. M. Warga and C. B. Kimmel, *Development* **108**, 569 (1990).
5. D. Strehlow and W. Gilbert, *Nature* **361**, 451 (1993).
6. H. Spemann and H. Mangold, *Wilhelm Roux Arch. Entwicklungsmech. Org.* **100**, 599 (1924); J. A. Oppenheimer, *J. Exp. Zool.* **72**, 409 (1936).
7. C. B. Kimmel and R. D. Law, *Dev. Biol.* **108**, 78 (1985).
8. Zebrafish (*Danio rerio*, Ekkwill Waterlife Resources, Gibsonton, FL) were maintained according to standard methods [M. Westerfield, Ed., *The Zebrafish Book* (Univ. of Oregon Press, Eugene, OR, 1993)]. Blastomeres were injected with fluorescein isothiocyanate–Dextran (2-megadalton molecular weight; Sigma) or Texas Red isothiocyanate–Dextran (2-megadalton molecular weight; Molecular Probes), prepared as described (*5*). Microinjection was performed under a stereomicroscope at a magnification of ×50. Dechorionated embryos were placed in 60-mm petri plates coated with 1% agarose in Instant Ocean salts (150 mg/liter). A V-shaped trough cut in the agarose was used to hold the embryos for microinjection. Embryos were transferred to agarose-coated dishes and placed at 28.5°C to develop. Injected embryos were screened with a Zeiss Axioplan microscope with epifluorescence optics within 20 min after microinjection and again 3 to 4 hours later. Embryos were discarded if (i) fluorescence was found in the yolk or in incorrect blastomeres, (ii) blastomeres contained both dyes, (iii) parts of labeled clones fluoresced with variable intensity, or (iv) embryos had abnormal morphology. In each injection batch, up to 20% of injected embryos were discarded. Injected embryos raised to 48 hours developed normally. Embryos were fixed in freshly prepared 4% paraformaldehyde (Polysciences, Niles, IL) in sucrose fix buffer (M. Westerfield, above).
9. S. E. Stachel, D. J. Grunwald, P. Z. Myers, *Development* **117**, 1261 (1993). Expression of *goosecoid* was detected by whole mount in situ hybridization as cited therein.
10. Embryos were oriented with the animal pole upward in 3% methyl cellulose. Fluorescence optics were used to observe the labeled clones; bright-field optics were used to observe *goosecoid* staining. Photographs of the embryos were traced so that the positions of the two labeled clones could be analyzed in relation to the dorsal midline. The position of the second cleavage plane was defined as the midpoint between the overlap of the two clones. The borders of each clone and the region of overlap could be viewed most clearly with fluorescence optics in the absence of white light. On average, the region of overlap covered an arc of 35°. The position of the dorsoventral axis was determined by drawing a line through the center of the gastrula and the midpoint of either the arc of *goosecoid* staining in early gastrula (40% epiboly) embryos or the center of the dot of *goosecoid* staining in midgastrula (60% epiboly) embryos. As expected, in half the embryos the region of overlap between the labeled clones fell on the ventral side of the embryo.
11. S. Abdelilah, L. Solnica-Krezel, D. Stainier, W. Driever, *Nature*, in press.
12. C. B. Kimmel and R. M. Warga, *Dev. Biol.* **124**, 269 (1987).
13. J. A. Oppenheimer, *J. Exp. Zool.* **73**, 405 (1936).
14. E. T. Wilson, K. A. Helde, D. J. Grunwald, *Trends Genet.* **9**, 348 (1993).
15. C. B. Kimmel, R. M. Warga, D. A. Kane, *Development* **120**, 265 (1994).
16. R. K. Ho and C. B. Kimmel, *Science* **261**, 109 (1993); S. Schulte-Merker, R. K. Ho, B. G. Herrmann, C. Nusslein-Volhard, *Development* **116**, 1021 (1992).
17. Embryos were injected as in Fig. 1D, allowed to develop to 60 to 75% epiboly, then fixed and mounted as described above. Observations were by Bio-Rad confocal laser scanning microscopy (MRC-600); projected *Z*-series were analyzed. Each embryo was oriented twice to show the central and marginal clones separately. The area occupied by each labeled clone was defined as the portion of the clone with pixel intensities within the range of 43 to 255, from the Histogram function of the software CoMOS 6.01 (Bio-Rad). For each embryo, the ratio (area of marginal clone/area of central clone) between the two areas was calculated. The mean of the ratios (n = 22) was determined after the most extreme high and low values were discarded. The spread of central clones was so extensive that only a portion could be captured in a single *Z*-series. In contrast, the vast majority of each marginal clone could be examined in a single orientation. As such, the ratio of the areas is an underestimate of the difference in scatter between the two types of clones.
18. Embryos were injected as described in Fig. 1D. At 60% epiboly, eight embryos were dissociated into single cells in Ca^{2+}-, Mg^{2+}-free phosphate buffered saline and fixed as described above. The dissociated cells of each embryo were treated with 4,6-diamine-2-phenylindole (DAPI) to label all nuclei. For each embryo, about 300 DAPI-positive cells were analyzed to identify cells also labeled with fluorescein or Texas Red.
19. Early cell division in the zebrafish is rapid and synchronous (*7*); later divisions are slower, asynchronous, and do not cluster in a particular region of the embryo [D. A. Kane and C. B. Kimmel, *Development* **119**, 447 (1993); D. A. Kane, R. M. Warga, C. B. Kimmel, *Nature* **360**, 735 (1992)].

20. We thank the members of our lab and the Department of Human Genetics for stimulating discussions during the course of these experiments as well as C. Thummel and P. Reid for their generosity and technical support. K.A.H. is a predoctoral trainee supported by a National Institutes of Health training grant. D.J.G. was supported by a Sloan Fellowship. Supported by grants from the University of Utah and the National Science Foundation (DCB-9105585).

13 May 1994; accepted 24 June 1994

The High-Resolution Crystal Structure of a Parallel-Stranded Guanine Tetraplex

Gerard Laughlan, Alastair I. H. Murchie, David G. Norman, Madeleine H. Moore, Peter C. E. Moody, David M. J. Lilley, Ben Luisi

Repeat tracts of guanine bases found in DNA and RNA can form tetraplex structures in the presence of a variety of monovalent cations. Evidence suggests that guanine tetraplexes assume important functions within chromosomal telomeres, immunoglobulin switch regions, and the human immunodeficiency virus genome. The structure of a parallel-stranded tetraplex formed by the hexanucleotide d(TG_4T) and stabilized by sodium cations was determined by x-ray crystallography to 1.2 angstroms resolution. Sharply resolved sodium cations were found between and within planes of hydrogen-bonded guanine quartets, and an ordered groove hydration was observed. Distinct intra- and intermolecular stacking arrangements were adopted by the guanine quartets. Thymine bases were exclusively involved in making extensive lattice contacts.

DNA and RNA containing runs of consecutive guanine bases may adopt four-stranded conformations based on the hydrogen-bonded guanine tetrad, or G quartet (*1–4*). These conformations may exist in a variety of isomeric forms, encompassing both parallel and antiparallel conformations, all of which are stabilized by monovalent ions such as sodium and potassium (*2*, *5*, *6*). Such tetraplexes have been implicated in a variety of biological roles, including the function of chromosome telomeres (*7*), the dimerization of the human immunodeficiency virus RNA genome (*8*), and the site-specific recombination of immunoglobulin genes (*1*), and their functional

G. Laughlan and B. Luisi, Medical Research Council Virology Unit, Church Street, University of Glasgow, Glasgow G11 5JR, UK.
A. I. H. Murchie, D. G. Norman, D. M. J. Lilley, Cancer Research Campaign, Nucleic Acid Structure Research Group, Department of Biochemistry, The University, Dundee DD1 4HN, UK.
M. H. Moore and P. C. E. Moody, Department of Chemistry, York University, Heslington, York YO1 5DD, UK.

importance is supported by the isolation of proteins that bind and promote the formation of tetraplex structures (9). Here, we have solved the crystal structure of a parallel-stranded tetraplex adopted by d(TGGGGT) in the presence of sodium ions to a resolution of 1.2 Å. The strands associate by the formation of hydrogen-bonded guanine tetrads, generating a right-handed helix containing four equivalent strands. The tetraplex is stabilized by the coordination of sodium ions, which can be found either between or within the planes of the guanine tetrads. In addition, we observed extensive ordered hydration in the four equivalent grooves and distinctive base-stacking contacts.

Table 1. Data quality and refinement summary. Diffraction data were collected from single crystals at 4°C with a Rigaku *R* axis IIC phosphorimager area detector. Data sets were collected from two crystals with monochromated radiation from a Cu (1.543 Å) source and from one crystal with a Mo (0.711 Å) source. The crystals belong to space group *P*1, with cell dimensions a = 28.76 Å, b = 35.47 Å and c = 56.77 Å and α = 74.39°, β = 77.64°, and γ = 89.73°. Images were integrated with the interactive program DENZO (*26*), and the data were merged, scaled, and processed with the CCP4 suite of programs (SERC Daresbury Laboratory). The structure was solved with molecular replacement and extensively refined (*27*). The final model, which has an overall deviation from ideality of 0.016 Å in bond lengths and 1.61° in bond angles, contains a total of 2453 nonhydrogen atoms, including 9 calcium ions, 14 sodium ions, and 514 water molecules.

Resolution shell lower limit (Å)	N_{obs}*		Percentage of completeness	*R* factor†	⟨*I*/σ⟩‡
	Total	Unique			
3.16	19,796	3,604	99.9	0.041	9.4
2.24	36,633	6,583	99.9	0.052	9.4
1.83	47,617	8,516	99.8	0.070	8.5
1.58	38,766	10,045	99.4	0.089	6.9
1.41	20,918	11,117	97.0	0.099	7.4
1.29	23,093	12,237	96.6	0.161	4.6
1.20	24,851	13,162	95.5	0.224	3.4
Total	211,674	65,264	97.7	0.056	7.0

*N_{obs} = number of diffraction maxima in the corresponding resolution range. †*R* factor = $\Sigma_h \Sigma_i | I(h)_i - \langle I(h)\rangle | / \Sigma_h \Sigma_i \langle I(h)\rangle$ where *h* represents unique reflection indices. ‡⟨*I*/σ⟩ = the average intensity/standard deviation ratio in each resolution range.

The hexanucleotide TG_4T crystallized as the anticipated four-stranded tetraplex in the presence of sodium ions (*10*, *11*). Four tetraplexes occupied the irreducible repeating unit of the crystal. We solved the structure by molecular replacement using a model that was generated on the basis of information provided by nuclear magnetic resonance (NMR) studies (*12*) and refined the model by simulated annealing and a water structure analysis procedure coupled with least squares refinement (Table 1). The resulting structure had an *R* factor of 12.4%. The strands of the tetraplexes form a right-handed helix and are aligned parallel to each other (Fig. 1). The cell is composed of two pairs of tetraplexes, where members of a pair stack coaxially and with opposite polarity (Fig. 1). The spacing between successive guanine tetrads within a tetraplex is uniform; furthermore, the same spacing was observed at the interface where stacked tetraplexes join, giving the impression of a continuous helix. Within the tetraplexes, the strands were found to associate by means of the expected hydrogen bonding between N1H and O6, and between N2H and N7, as first proposed by Gellert *et al.* (*13*) (Fig. 1D). These tetrads exhibit a high degree of planarity and are connected to the backbones by glycosyl bonds that are exclusively anti; this is in contrast to the regular alternation of syn and anti glycosyl bonds in antiparallel gua-

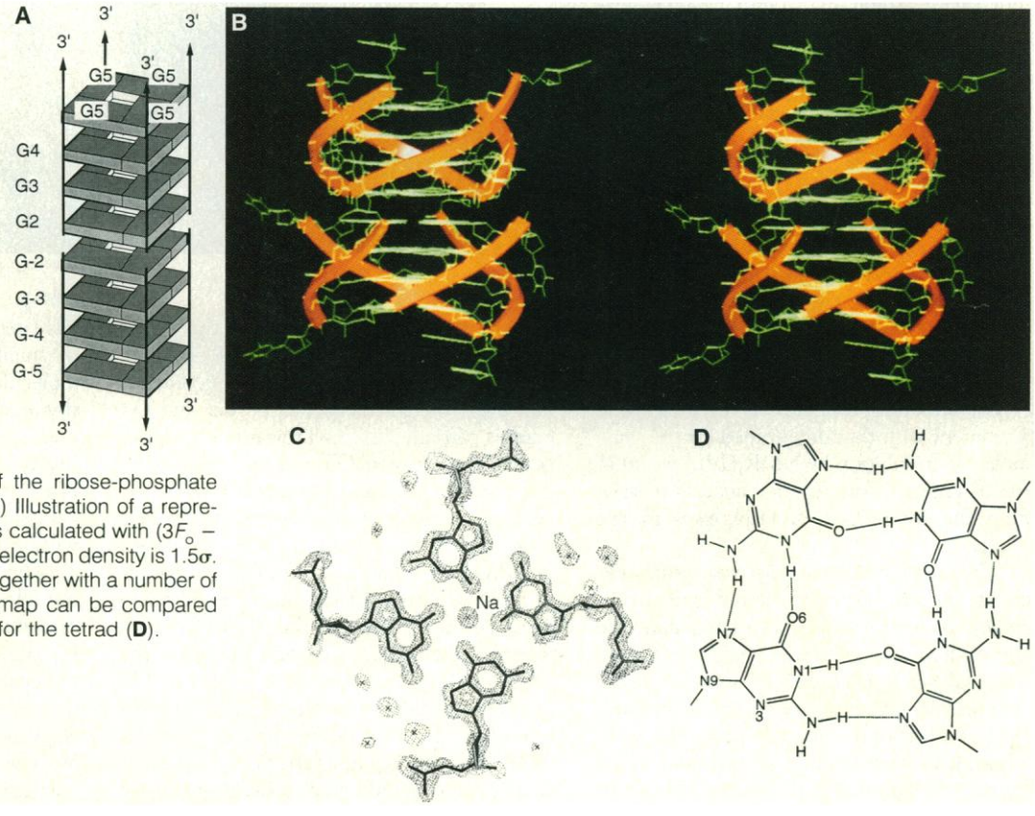

Fig. 1. Relative orientation and packing of tetraplexes in the crystal, and the structure of the guanine tetrad. (**A**) Schematic representation of the strand orientation of the DNA tetraplex and packing in the cell. Four parallel-stranded tetraplexes occupy the asymmetric unit. Tetraplex pairs stack with opposite polarity, with the interface formed by their 5′ ends. A second stacked pair (not shown here) is related to the first by a noncrystallographic translation. The numbering of the guanine tetrad planes is derived from 5′-T(1)G(2)G(3)G(4)G(5)T(6)-3′. Bases of the lower tetrad have been given negative integers for clarity. (**B**) A stereoscopic view of one stacked pair of tetraplexes, with the paths of the ribose-phosphate backbones highlighted by ribbons. (**C**) Illustration of a representative guanine tetrad. The map was calculated with $(3F_o - 2F_c)$ coefficients. The contouring of the electron density is 1.5σ. The central sodium ion can be seen, together with a number of water molecules in the grooves. The map can be compared with the proposed chemical structure for the tetrad (**D**).

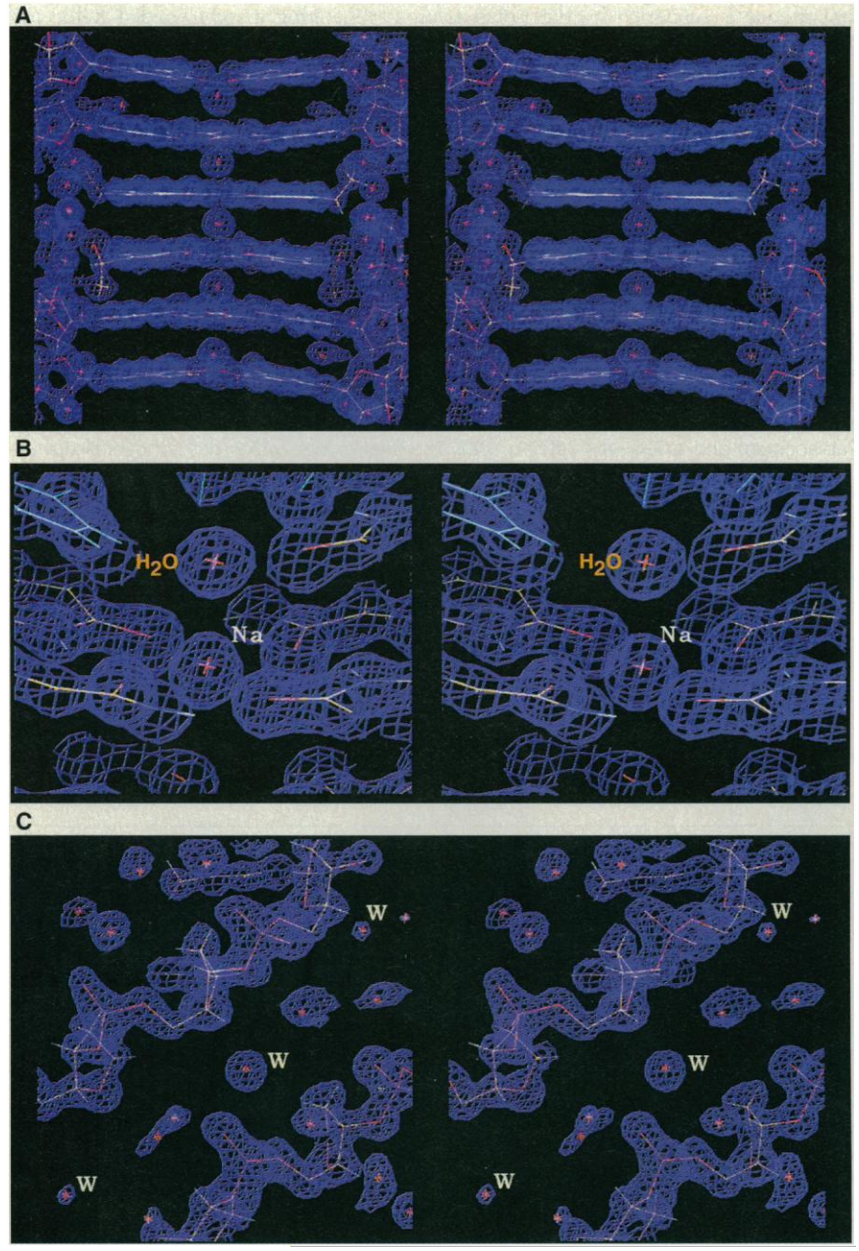

Fig. 2. Location of metal ions and water molecules. (**A**) The intertetrad sodium ions. Shown is a ($3F_o - 2F_c$) map at a contouring of 1.5σ, depicting the string of sodium ions running up the central cavity of the tetraplex. Each sodium ion (except at G5 and at G−5) is coordinated by eight guanine O6 atoms from the tetrad planes above and below. Only the central sodium located at the interface between tetrad planes is equidistantly coordinated by eight carbonyl oxygen atoms, with an average value for the two independent sites of 2.75 ± 0.02 Å. The ions coordinated between tetrad planes G2 and G3 are shifted toward the G3 plane with an average value for four independent sites of 2.89 ± 0.07 Å and 2.60 ± 0.07 Å for the G2 and G3 distances, respectively. The displacement is more pronounced for the ion located between the G3 and G4 planes, where the average distances are 3.32 ± 0.10 Å and 2.43 ± 0.03 Å for the G3 and G4 distances, respectively. The distances between G2 and G3 and between G3 and G4 pairs were averaged over four independent examples occurring within the asymmetric unit. Finally, there is no ion situated between the G4 and G5 planes because the sodium ions are located coplanar with the G5 tetrads. (**B**) The intratetrad sodium ions, occupying the tetrad planes at G5 and G−5. There are four independent examples of the sodium ion coplanar with the G5 tetrad. Each ion is coordinated by four guanine O6 atoms, with an average distance of 2.34 ± 0.02 Å, and by a water molecule at the axial position with an average distance of 2.53 ± 0.03 Å (the error represents the standard deviation of the mean). (**C**) The spine of hydration in the groove. W, water.

nine tetraplex structures (*14*, *15*).

The four independent tetraplexes are not entirely equivalent structurally, and the differences occur principally at the interface between coaxial tetraplexes. Most of the guanine sugars exhibit a B DNA character, having a pseudorotation angle of 158.5° ± 11.3° and an average phosphate separation of 6.63 ± 0.14 Å. However, at the tetraplex interface all of the G2 sugars of one of the coaxial tetraplex pairs switch to a conformation resembling that of A DNA (for example, with an averaged pseudorotation angle of 45.6° ± 8° and a phosphate separation of 6.93° ± 0.15 Å). This switch is apparently required to avoid steric clash at the tightly packed interface. Elsewhere in the molecule, where the sugars are uniformly like B DNA, the helical structure has a repeat corresponding to 10.4 tetrads per turn. This conformation of the crystal structure tetraplex is in reasonable agreement with that determined for the same molecule in solution by NMR (*12*), except at the 5′ end; this discrepancy appears to arise from the interaction of tetraplexes in the crystal.

The electron density map was of sufficient quality at early stages of the refinement procedure to identify sodium ions lying along the axis of the tetraplex. Seven sodium ions were observed per stacked pair of tetraplexes, including one located at the tetraplex interface (Fig. 2). Most of the ions are located between consecutive tetrads and are coordinated by the tetrad's eight buried O6 atoms. The sodium ion that is located between the tetrad planes of G2 and G−2 satisfies a nearly perfect bipyramidal coordination geometry. However, the ion between the G2 and G3 bases is partially displaced toward G3, whereas the ion between G3 and G4 exhibits an even greater displacement toward G4. This general outward displacement of sodium ions away from the central pair G2 and G−2 tetrad step is probably the result of electrostatic repulsion between cations. The displacement continues until at the 3′ end of each tetraplex the sodium ions actually lie in the plane of the tetrad formed by the four G5 bases. This coplanar sodium ion is coordinated by the four O6 atoms and by a water molecule that is bound at the fifth position. A similar pentacoordinated sodium geometry has been observed in crystallographic studies of small molecules (*16*). The axial water molecule forms hydrogen bonds with two thymine bases from a symmetry-related molecule and another water molecule to satisfy a nearly perfect tetrahedral geometry. Because of its larger ionic radius, it is unlikely that potassium could adopt a position coplanar with a tetrad.

The parallel orientation of strands in the tetraplex generates four equivalent grooves of essentially minor groove character, containing the N2, N3, and C8 atoms. The edge of the guanine bases that would define the major groove in duplex DNA lies on the interior of the tetraplex. The relatively narrow grooves are favorable binding sites for water molecules, which form hydrogen bonds with the

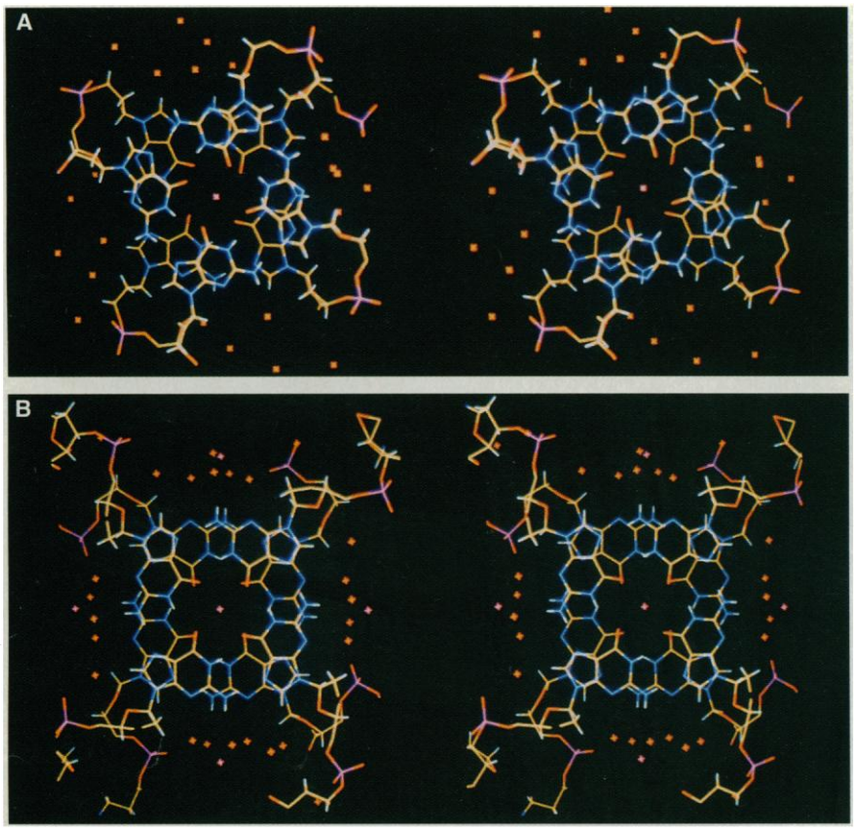

Fig. 3. Stacking between tetrads. (**A**) Guanine tetrad stacking within one tetraplex, at tetrads G3 and G4. (**B**) Stacking at the interface between tetraplexes at tetrads G2 and G−2.

exposed N2 amino groups, the heterocyclic N3, and each other, forming a well-organized hydration network (Fig. 2C).

At the tetraplex interface where G2 and G−2 meet, there is extensive overlap of the five-membered rings of the guanine bases, but the six-membered rings have no effective overlap (Fig. 3). This stacking is not possible in a regular DNA duplex with Watson-Crick base pairs; however, an analogous interaction occurs at positions where the glycosyl bond changes from the syn to the anti conformation in antiparallel tetraplex structures (*15*). That the tetraplexes studied here selected this distinctive stacking conformation might suggest that it is energetically preferred. For all other tetrad steps of TG_4T, the adjacent guanine bases stack such that the N2 of one is sandwiched between the N7 and N3 of the tetrads above and below (Fig. 3). As the protons of N2 are electropositive but N7 and N3 are electronegative, it seems likely that the favorable overlap of partial charges contributes to the stability of the stacking conformation.

The high crystalline order of the tetraplex may be attributed to favorable lattice contacts. In addition to the 5′ end-to-end stacking of the tetraplex pairs within the cell, there are also longitudinal contacts with molecules in the neighboring cell, resulting in the formation of a pseudocontinuous helix that runs along the length of the crystal. The tetraplexes in adjacent crystal cells interdigitate such that the thymine bases at the 3′ ends form complementary T·T base pairs. In addition, there are lateral interactions between the tetraplexes that are made by the stacking of thymine bases originating from both the 3′ and 5′ termini of several molecules. Some thymine bases are conformationally disordered. A well-defined thymine tetrad of the kind formed by uracil bases in $r(UG_4U)$ (*17*) was not observed in this crystalline DNA structure, which is in agreement with NMR studies in solution (*12*, *18*). Calcium ions were included in the crystallization solution, and nine ions are well ordered, each with a well-defined hydration sphere. The ions do not directly contact the phosphate groups of the DNA, as expected, but interact by means of the water molecules.

This tetraplex structure has been solved to one of the highest resolutions ever achieved for a nucleic acid. This species is parallel-stranded, but the information on metal ion coordination and tetrad-tetrad interaction that is obtained at this resolution may provide insight that is relevant to all guanine tetraplexes, including both parallel and antiparallel foldback structures. Such structural detail is likely to be important in the analysis of tetraplex-protein interactions and thus in understanding the biological function of these structures.

REFERENCES AND NOTES

1. D. Sen and W. Gilbert, *Nature* **334**, 364 (1988).
2. J. R. Williamson *et al.*, *Cell* **59**, 871 (1989).
3. W. I. Sundquist and A. Klug, *Nature* **342**, 825 (1989).
4. J. Kim *et al.*, *ibid.* **351**, 331 (1991).
5. D. Sen and W. Gilbert, *ibid.* **344**, 410 (1990).
6. C. C. Hardin, E. Henderson, T. Watson, J. K. Prosser, *Biochemistry* **30**, 4460 (1991).
7. E. H. Blackburn, *Nature* **350**, 569 (1991).
8. W. I. Sundquist and S. Heaphy, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3393 (1993).
9. G. W. Fang and T. R. Cech, *Cell* **74**, 875 (1993).
10. The oligonucleotide 5′-TpGpGpGpGpT-3′ was synthesized on a scale of 1 μM with use of β-cyanoethylphosphoramidite chemistry [S. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.* **22**, 1859 (1981); N. D. Sinha, J. Biernat, J. McManus, H. Koster, *Nucleic Acids Res.* **12**, 4539 (1984)] without 5′ deprotection and purified by reversed-phase chromatography, followed by detritylation. Tetraplex formation was induced by slow cooling from 70°C in 5 mM Hepes (pH 7.0) and 0.1 M NaCl and then dialyzed against 50 mM NaCl with the use of a 3-kD molecular weight cutoff membrane (centricon-3, Amicon, Beverly, MA). Mass spectrometry [U. Pieles, W. Zürcher, M. Schar, H. E. Moser, *Nucleic Acids Res.* **21**, 3191 (1993)] of this material revealed a single peak with the expected molecular mass (1862.4 daltons).
11. Crystals were grown by vapor diffusion at 4°C with the use of hanging drops. DNA was quantified spectrophotometrically with extinction coefficients calculated from the base composition. Crystals used for data collection were grown from hanging drops containing 20 mM Na cacodylate·HCl (pH 6.6), 12 mM $CaCl_2$, 6 mM spermine, 130 to 180 mM NaCl, 5% v/v methylpentanediol (MPD), and 1 mM TG_4T. The drops were equilibrated against wells containing 120 mM sodium cacodylate·HCl (pH 6.6), 70 mM $CaCl_2$, 700 mM to 1 M NaCl, and 26 to 32% MPD. Crystals (which were usually twinned) grew within 2 weeks to a size of about 340 μm by 200 μm by 200 μm. The crystals were stable in the liquor of the hanging drops and were mounted directly after dissection into individual components.
12. F. Aboul-ela *et al.*, *Nature* **360**, 280 (1992); F. Aboul-ela, A. I. H. Murchie, D. G. Norman, D. M. J. Lilley, *J. Mol. Biol.*, in press.
13. M. Gellert, M. N. Lipsett, D. R. Davies, *Proc. Natl. Acad. Sci. U.S.A.* **48**, 2013 (1962).
14. Y. Wang *et al.*, *Nucleic Acids Res.* **19**, 4619 (1991); Y. Wang *et al.*, *J. Mol. Biol.* **222**, 819 (1991); C. Kang *et al.*, *Nature* **356**, 126 (1992); Y. Wang and D. J. Patel, *Biochemistry* **31**, 8112 (1992); *Structure* **1**, 263 (1993).
15. F. W. Smith and J. Feigon, *Biochemistry* **32**, 8682 (1993).
16. I. L. Karle, *ibid.* **13**, 2155 (1974).
17. C. Cheong and P. B. Moore, *ibid.* **31**, 8406 (1992).
18. Y. Wang and D. J. Patel, *J. Mol. Biol.* **234**, 1171 (1993).
19. J. Navaza, *Acta Crystallogr.* **A46**, 619 (1990).
20. The Patterson function revealed a strong peak at nearly (0, 0, 1/2), arising from a noncrystallographic translation parallel with the *c* axis. This translation was also manifested as a systematic attenuation of intensities for reflections with odd-numbered *l* indices. A noncrystallographic fourfold rotation axis was also found, approximately 11° off the *b* axis, and coincident with the axis defining the orientation of a strong base-stacking transform.
21. A. T. Brünger, *X-PLOR: A System for X-Ray Crystallography and NMR* (Yale Univ. Press, New Haven, 1993).
22. V. S. Lamzin and K. S. Wilson, *Acta Crystallogr.* **D49**, 129 (1993).
23. W. A. Hendrickson and J. H. Konnert, in *Biomolecular Structure, Function, Conformation and Evolution*, R. Srinivisan, Ed. (Pergamon, Oxford, 1980), vol. 1, pp. 43–57.

24. T. A. Jones, *J. Appl. Crystallogr.* 11, 272 (1978).
25. SHELX-93; G. Sheldrick, University of Göttingen.
26. Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend*, L. Sawyer, N. Issacs, S. Bailey, Eds. (Science and Engineering Research Council, Daresbury Laboratory, 1993), pp. 56–62.
27. A number of parallel-stranded guanine tetraplex models were generated on the basis of the conformational details determined in solution by NMR (*12*). These were then incrementally over- and underwound to generate a series of model structures for molecular replacement with AMORE (*19*). The optimal model yielded a set of highest peaks related by noncrystallographic fourfold symmetry. Translation function solutions were found that were consistent with the noncrystallographic symmetry suggested by the Patterson function (*20*). These were initially refined by simulated annealing with X-PLOR (version 3.1) (*21*), and only restraints to ensure base planarity were used. An automated refinement procedure followed (ARP) (*22*) that reiteratively adjusts the water structure of the model. The resulting map was sufficiently clear to reveal that the derived solution was false, in that one of the tetraplexes in each stacked pair was spatially out of register by one base step along the helical axis. The error was manually rectified, and the genuine solution rapidly refined with X-PLOR and ARP, combined with least squares minimization [PROTIN/PROLSQ (*23*)] with the use of a modified stereochemical dictionary to an *R* factor of 18.8%. Extensive rebuilding was undertaken throughout with the use of FRODO (*24*), resulting in the incorporation of 27 of the 32 thymine bases. Further refinement, including anisotropic temperature factors and the introduction of hydrogen atoms, was then undertaken using SHELX-93 (*25*) to a final *R* factor of 12.4% for the data from 8.0 to 1.2 Å. The free *R* factor (*21*) was 17.6% for 1 out of 10 of the reflections being excluded from the refinement procedure. Even after anisotropic refinement, there was no clear density for the five missing thymine bases, presumably because of high thermal disorder, and these were excluded from the final model.
28. We are grateful to J. Navaza for help with AMORE; U. Pieles for mass spectrometry; E. Dodson, A. Lapthorn, F. Aboul-ela, and Z. Otwinowski for helpful suggestions; N. Isaacs for support; and the Cancer Research Campaign and Medical Research Council for financial assistance. The coordinates will be deposited with the Brookhaven Data Bank and can be obtained immediately by contacting B.L. (ben@chem.glasgow.ac.uk).

7 March 1994; accepted 17 May 1994

The Three-Dimensional Crystal Structure of the Catalytic Core of Cellobiohydrolase I from *Trichoderma reesei*

Christina Divne, Jerry Ståhlberg, Tapani Reinikainen, Laura Ruohonen, Göran Pettersson, Jonathan K. C. Knowles, Tuula T. Teeri, T. Alwyn Jones*

Cellulose is the major polysaccharide of plants where it plays a predominantly structural role. A variety of highly specialized microorganisms have evolved to produce enzymes that either synergistically or in complexes can carry out the complete hydrolysis of cellulose. The structure of the major cellobiohydrolase, CBHI, of the potent cellulolytic fungus *Trichoderma reesei* has been determined and refined to 1.8 angstrom resolution. The molecule contains a 40 angstrom long active site tunnel that may account for many of the previously poorly understood macroscopic properties of the enzyme and its interaction with solid cellulose. The active site residues were identified by solving the structure of the enzyme complexed with an oligosaccharide, *o*-iodobenzyl-1-thio-β-cellobioside. The three-dimensional structure is very similar to a family of bacterial β-glucanases with the main-chain topology of the plant legume lectins.

Cellulose is the most abundant renewable resource on Earth, accounting for about half of the organic material in the biosphere. In addition to its ecological and commercial importance, the high degree of crystallinity of the substrate makes cellulose degradation a problem of fundamental interest. Cellulolytic fungi produce exoglucanases or cellobiohydrolases (CBH, 1,4-β-D-glucan cellobiohydrolase, E.C. 3.2.1.91) that hydrolyze crystalline cellulose by cleavage from chain ends (*1*). They act synergistically with endoglucanases (EG, 1,4-β-D-glucanohydrolase, E.C. 3.2.1.4) that primarily hydrolyze the disordered, amorphous regions of cellulose, cutting at internal glycosidic bonds (*1*).

C. Divne, J. Ståhlberg, T. A. Jones, Department of Molecular Biology, Uppsala University, Biomedical Centre, Box 590, S-751 24 Uppsala, Sweden.
T. Reinikainen, L. Ruohonen, T. T. Teeri, VTT Biotechnology and Food Research, Post Office Box 1500, FIN-02044 VTT, Espoo, Finland.
G. Pettersson, Department of Biochemistry, Uppsala University, Biomedical Centre, Box 576, S-751 23 Uppsala, Sweden.
J. K. C. Knowles, Glaxo Institute for Molecular Biology, 14 Chemin des Aulx, 1228 Plan-les-Ouates, Geneva, Switzerland.

*To whom correspondence should be addressed.

The cellobiohydrolase CBHI is probably the key enzyme needed for the efficient hydrolysis of native crystalline cellulose. It is the most abundant cellulase produced by the filamentous fungus *Trichoderma reesei*, and the removal of its gene reduces overall activity on crystalline cellulose by 70% (*2*). It exhibits the strongest synergy with other *T. reesei* and bacterial cellulases (*3*). Because CBHI acts synergistically with the other cellobiohydrolase (CBHII) produced by *T. reesei* (*4*), it has been suggested that one or both of the enzymes may possess endoglucanase activity (*5*). Cellobiohydrolase I has been classified on the basis of its amino acid sequence as a family C enzyme (*6*). This family includes both exo- and endoglucanases which cleave the β(1 → 4) glycosidic bond by a double-displacement mechanism, resulting in retention of configuration of the product, cellobiose (*7*).

Cellobiohydrolase I is a two-domain enzyme, consisting of a large catalytic core linked to a small cellulose-binding domain by a heavily glycosylated linker region (*8*). The structure of the core was solved by the standard method of multiple isomorphous

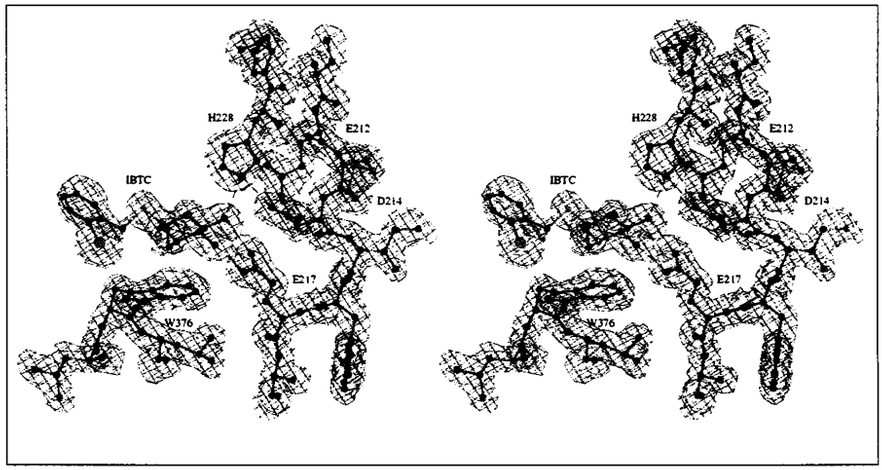

Fig. 1. Electron density of the active site residues. The map was calculated with $2|F_{obs}| - |F_{calc}|$ amplitudes, and phases were calculated from the final model. The ligand, *o*-iodobenzyl-1-thio-β-D-glucose, is labeled IBTC and hydrogen bonds are indicated by dashed lines.