Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 867–870, February 1990
Biochemistry

G-DNA: A twice-folded DNA structure adopted by single-stranded oligo(dG) and its implications for telomeres

[DNA methylation/quadruple helix/DNA loops/computer simulation/chromosome-end $(dG)_n$ sequences]

I. G. PANYUTIN*, O. I. KOVALSKY*, E. I. BUDOWSKY*, R. E. DICKERSON†, M. E. RIKHIREV‡, AND A. A. LIPANOV‡§

*N. D. Zelinsky Institute of Organic Chemistry, U.S.S.R. Academy of Sciences, 47 Leninsky Prospekt, Moscow 117913, U.S.S.R.; †Molecular Biology Institute, University of California at Los Angeles, Los Angeles, CA 90024-1570; and ‡Institute of Molecular Genetics, U.S.S.R. Academy of Sciences, 46 Kurchatov Square, Moscow 123182, U.S.S.R.

Contributed by R. E. Dickerson, November 7, 1989

ABSTRACT **Our dimethyl sulfate modification experiments suggest that $(dG)_n$ stretches within single-stranded DNA fragments, which represent the simplest model for telomeric sequences, adopt a complex intrastrand structure other than a simple hairpin. We present a molecular model for the DNA structure that conforms to dimethyl sulfate methylation data. The principal element of this G-DNA structure is a quadruple helix formed by pairwise antiparallel segments of the twice-folded $(dG)_n$ stretch. This quadruple core has two wide and two narrow grooves connected by three loop-shaped segments. The strong stacking interactions of the neighboring guanine tetrads and the large number of hydrogen bonds formed can be the primary reasons that such structures are favored over a common hairpin for long $(dG)_n$ stretches. Such compact structures may be formed from $(dG)_n$ stretches of telomeric sequences.**

A number of unusual DNA structures (cruciform, Z form, H form) have been shown to form within a variety of specific sequences (for reviews, see refs. 1–3) and at the present time are being investigated for biological relevance in several laboratories. Under normal conditions the B form is the most favorable thermodynamically (4–6). Transitions to alternative conformations require specific external conditions, supercoiling being among the most physiological. Hence, it is not surprising that most alternative DNA forms have been investigated in closed circular DNAs under superhelical stress. An entirely different situation is realized at the ends of chromosomes, where a special enzyme—telomerase (7, 8)—extends the 3′ end by adding $(dG)_n$-containing sequences repeated many times. These single-stranded ends are reputed to be capable of forming hairpin-like structures stabilized by hydrogen bonds between guanines (9). These structures are believed to play the leading role in chromosome-end replication models (10).

The simplest model for DNA enriched in $(dG)_n$ regions is simply oligo(dG). We have constructed plasmids pPG37 and pPG27, which contain $(dG)_n$·$(dC)_n$ inserts at the *Pst* I site of pUC18 with $n = 37$ and 27, respectively. It has been found (11) that, after the short insert-containing restriction fragments have been denatured, the $(dG)_n$- and $(dC)_n$-containing strands renature much more slowly than do those of arbitrary sequence. In addition, the $(dG)_n$ strand has an abnormally high mobility in polyacrylamide gel electrophoresis (PAGE) (11). These data argue in favor of the existence within $(dG)_n$-containing single-stranded fragments of an intrastrand structure (G structure) stabilized by formation of hydrogen bonds between guanines. In this paper we describe the results of methylation experiments on the oligo(dG)-containing fragments, and present a computer model that explains the unusual methylation pattern obtained for this single-stranded DNA.

MATERIALS AND METHODS

Sample preparation and details of the experiments are described at length elsewhere (11) and will only be summarized here. Plasmids pPG27 and pPG37 were cleaved with *Eco*RI, 3′-end-labeled with the Klenow fragment of DNA polymerase I, and cleaved with *Hin*dIII, yielding shorter $(dG)_n$-containing fragments of length $52 + n$ base pairs, which were isolated by preparative gel electrophoresis. The fragments with the inserts (about 2 ng in 10 μl) were denatured in 10 mM Tris·HCl, pH 8.0/10 mM NaCl/1 mM EDTA by incubation at 100°C for 1 min. Samples were quickly chilled and loaded on a nondenaturing polyacrylamide gel. The band corresponding to the $(dG)_n$-containing strand was extracted from the gel and methylated by the method of Maxam and Gilbert (12). Samples were analyzed on an 8% polyacrylamide gel containing 8.3 M urea at 60°C. Autoradiographs were scanned with an Ultroscan densitometer (LKB) conjugated with an IBM PC/XT via a GS software package (U.S.S.R.).

RESULTS AND DISCUSSION

Methylation Data. The simplest folded model for the poly(dG) chain would be a hairpin stabilized by Hoogsteen-like G·G pairs. (See ref. 13 for a review of base-pairing schemes.) In this case, one $(dG)_n$ arm of the hairpin should be protected against dimethyl sulfate methylation at its N7 positions because of base pairing. This is exactly what was observed in sequencing a single-stranded $(dG)_{23}$ stretch from the sea crab genome (14). However, the G-structure methylation pattern for longer stretches—$(dG)_{27}$ and $(dG)_{37}$—is more complicated (Fig. 1). Clear maxima and minima of modification are seen against the background of general diminution of peak amplitude associated with overmodification (more than one dimethyl sulfate modification per fragment). The maxima between extremities are positioned at one-quarter, one-half, and three-quarters of the way along the insert. This suggests that the $(dG)_n$ strand is bent in three places, with three exposed and methylated loops. The fact that all four segments flanking the loops are simultaneously protected against methylation at their N7 positions can only mean (15) that these loci are shielded by formation of Hoogsteen-like hydrogen-bonded tetrads (16, 17) like that shown in Fig. 2. But in any event, the similarity of disposition of maxima and minima for the $(dG)_{27}$ and $(dG)_{37}$ oligomers shows that these two share a common structure, whatever

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

§Present address: Molecular Biology Institute, University of California at Los Angeles, Los Angeles, CA 90024-1570.

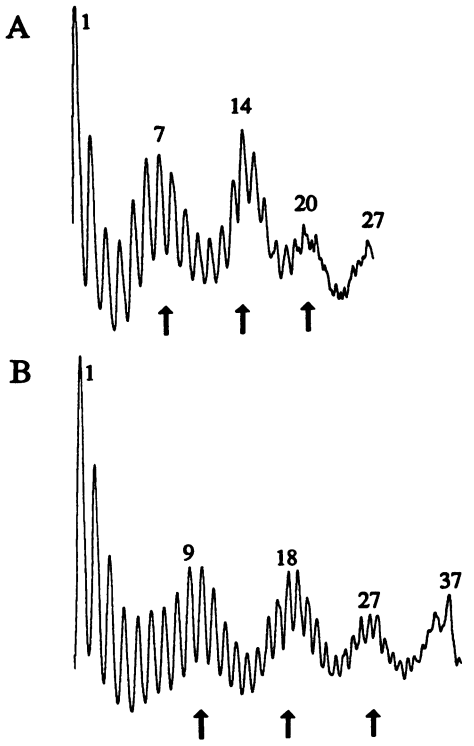

FIG. 1. Distribution of guanine methylation within the $(dG)_{27}$ (*A*) and the $(dG)_{37}$ (*B*) inserts, which form a complex intramolecular G structure within single-stranded restriction fragments. Nucleotide numbers are shown. According to our model, the three peaks (marked by arrows) between the extremities, corresponding to terminal guanines, correspond to three loops in the structure. Electrophoretic experiments with labeled restriction fragments carrying the $(dG)_n$ stretches (11) clearly indicate the absence of multistranded complex formation, since these fragments migrate in PAGE much faster than ones carrying $(dC)_n$ stretches or the native duplexes carrying $(dG)_n \cdot (dC)_n$.

that might be—a structure that is different than that for $(dG)_{23}$.

Molecular Model of G-DNA. One simple way of achieving this pattern of four protected stem regions and three exposed loops is illustrated in Fig. 3. The simple hairpin now is doubled back on itself to build a fourfold helix, with adjacent chains around the helix running in opposite directions. We

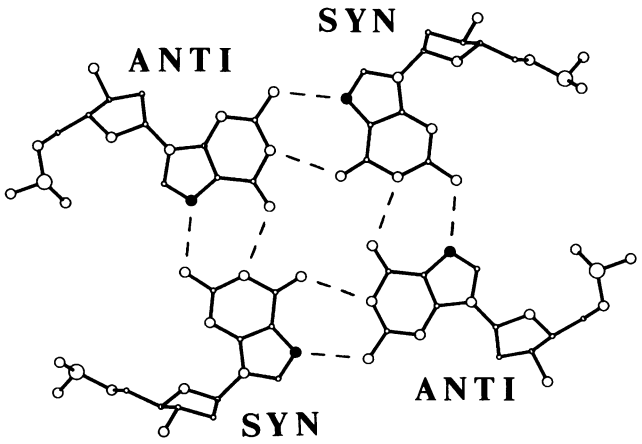

FIG. 2. A G_4 tetrad with Hoogsteen-like hydrogen bonding between guanines. Dashed lines show hydrogen bonds. N7 atoms are in black. Glycosyl bonds are shown in the alternating syn–anti conformations that would be expected in hairpin structures (9, 18). Hence, the fourfold symmetry of the bases is reduced to twofold symmetry overall.

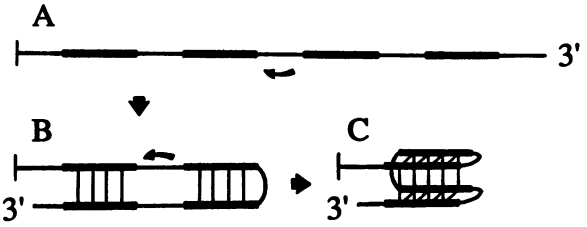

FIG. 3. A straight deoxyoligonucleotide chain with regions of oligo(dG) (*A*) can be folded back into a hairpin (*B*), and the hairpin can be folded once more to build a four-chain alternating antiparallel helix (*C*). We propose similar structures for a telomeric sequence. (*A*) The single-stranded end after telomerase grafting [$(dG)_n$ stretches are shaded]. (*B*) Folding into a hairpin with Hoogsteen-like G·G pairs. (*C*) Convolution into a compact structure with formation of quadruple helix between the $(dG)_n$ stretches.

have constructed a molecular model based on this folded hairpin for a $(dG)_{27}$ oligomer (Fig. 4). The core of the model contains 16 guanines forming a helix of four tetrads (Fig. 2), and each of the three loops 1–3 contains three unpaired guanines. The base tetrad itself possesses fourfold rotational symmetry, but this is degraded to twofold symmetry by the sugar–phosphate backbone. The strands of that backbone (Fig. 4*A*) run in opposite directions past any two adjacent hydrogen-bonded guanines in a tetrad in accordance with the folded-hairpin model.

This antiparallel sense of adjacent chains around the fourfold helix requires that the glycosyl bonds of adjacent chains have opposite conformations: syn vs. anti, in long helices. To see this, consider in isolation only the bottom two bases of the tetrad drawn in Fig. 2, and imagine that the lower left guanine were rotated 180° about its glycosyl bond to make a purine–purine Watson–Crick-like base pair. The conformation then would be anti about both strands of this double helix, as is required in any antiparallel double helix whose backbone is not reversed intermittently as in Z-DNA. This symmetry argument is also corroborated by calculations (18, 20) and by NMR evidence on real hairpins (9).

The sugar–phosphate chains adjacent to the guanine core were constructed via computer simulation in a manner that avoided overlap of nonbonded atoms. Since each of the four coupled bases lies significantly off the helix axis, the conformation of the chains resembles A-form DNA rather than B (21). The three loop segments were generated by small variations in the torsional angles, starting from the conformation of the rest of the chain. To improve the internal energy, the structure thus obtained was put through the X-PLOR molecular dynamics refinement programs (22), which simulate heating the structure and cooling again with subsequent minimization. This treatment produced the final model presented in Fig. 4, in which all van der Waals contacts and hydrogen bond lengths are acceptable.

The alternation of syn and anti glycosyl bonds in the quadruple core of the structure produces two narrow and two wide grooves. The narrow grooves are capped by loops 1 and 3 at the bottom of Fig. 4 *B* and *C*, and the wide grooves are capped by loop 2 and by the two free ends of the chain at the top of each part of Fig. 4. It is possible that the unpaired bases of loops 1 and 3 can form short-lived hydrogen-bonded interactions, furnishing a small additional measure of methylation protection that could account for the central methylation maximum in Fig. 1 being relatively higher than the two neighboring maxima (taking into account the overall drop in peak intensity from left to right).

Earlier proposals for quadruple helices based on G·G·G·G tetrads considered only parallel chains (15–17) because only in this manner could all glycosyl bonds be kept in the energetically favored anti form. This advantage must be

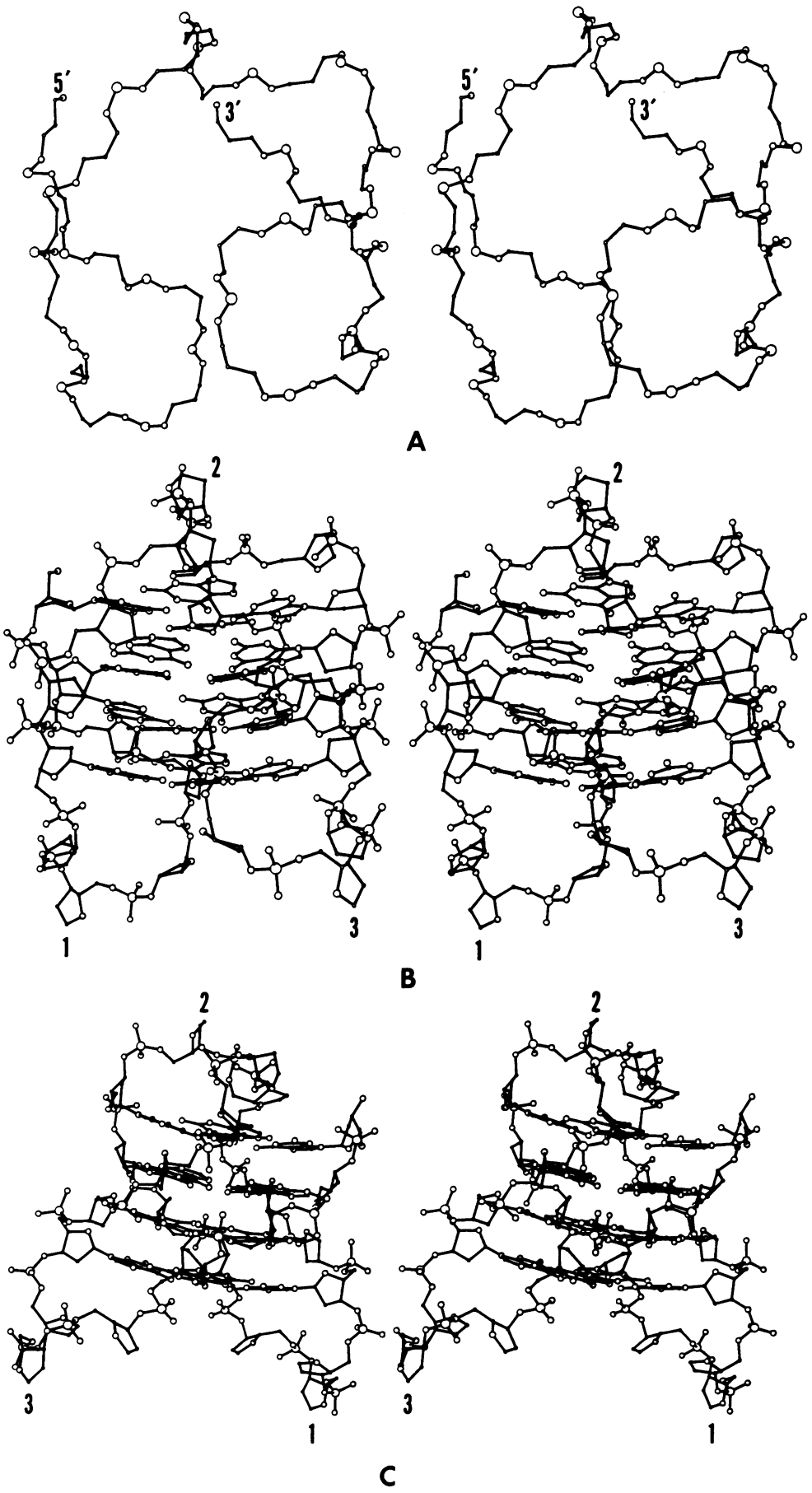

FIG. 4. The proposed structure for $(dG)_{27}$ with a fourfold helix core. (*A*) Backbone chain only, illustrating folding analogous to that in Fig. 3*C*. (*B*) Full G-DNA structure, with four stacked G_4 tetrads, in a view into one of the two wide grooves. The three external loops are numbered. Unpaired guanine bases in these loops are omitted for clarity. The 16-guanine-base core is consistent with fiber diffraction data for tetrads with parallel chains (19); these bases were held fixed during computer optimization. Twofold symmetry also was maintained for the structure with the exception of central loop 2. One unpaired guanosine should be added to each end of the chain to obtain a model for the full $(dG)_{27}$. (*C*) View 90° to the right, looking into one of the two narrow grooves. The width of this groove is quite sensitive to the particular values of electrostatic terms chosen for energy refinement.

abandoned if one is to build the helix from once-folded hairpins or twice-folded tetra-chain structures. This folded structure is favored over the common hairpin for the $(dG)_{27}$ oligomer because of strong stacking interactions between

neighboring G_4 tetrads, and because of the increased number of hydrogen bonds: 32 as compared with only 22 in a hairpin. The extra stabilizing contribution of these factors will increase with the length of the $(dG)_n$ stretch. One destabilizing factor is the number of unpaired bases in the loops: nine in three loops for the tetrad vs. about three in one loop in the hairpin, but the significance of this decreases with longer $(dG)_n$ stretches.

Implications for Telomeres. It should be noted that G-structure formation in a long single-stranded $(dG)_n$ stretch in solution does not require any extra stabilizing factors: acid pH, high salt, etc. Apparently, under certain conditions this structure is also typical of a series of other $(dG)_n$-containing sequences. Guanine-rich DNA segments are frequently found in eukaryotic genomes. Such segments with regularly repeated $(dG)_n$ stretches have been found in all telomeres studied so far (23, 24). Fig. 3 shows a plausible folding of a telomere sequence like poly(G_n·N_m), where N is any nucleoside. One can speculate about the possible biological function of such structures. For example, because of the inability of a folded G-DNA region of oligomer to form hydrogen bonds with another strand, G-DNA might help to control recombination of chromosome ends. It is also possible that studying the formation of a structure like this will be helpful in creating the best model for the replication of telomeres.

Models such as this are valuable as stimuli to further research. An oligomer containing $(dG)_n$ stretches has recently been synthesized at the University of California, Los Angeles for x-ray and NMR studies in collaboration with Juli Feigon and Kazimierz Grzeskowiak.

Note Added in Proof. While this paper was in press, two additional lines of evidence have appeared that suggest formation of a similar antiparallel quadruple helix within telomeric sequences of *Oxytricha* and *Tetrahymena* (25, 26). These results support our conclusion that the G-DNA structure adopted by oligo(dG) also is typical of a series of other $(dG)_n$-containing sequences, including biologically significant ones.

We thank Mary Kopka for useful discussions and Kazunori Yanagi for consultations regarding refinement.

1. Rich, A., Nordheim, A. & Wang, A. H.-J. (1984) *Annu. Rev. Biochem.* **53,** 791–846.
2. Wells, R. D. (1988) *J. Biol. Chem.* **263,** 1095–1098.
3. Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M. & Wohlrab, F. (1988) *FASEB J.* **2,** 2939–2949.
4. Peck, L. T. & Wang, J. C. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6206–6210.
5. Frank-Kamenetskii, M. D. & Vologodskii, A. V. (1984) *Nature (London)* **307,** 481–482.
6. Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. (1986) *J. Biomol. Struct. Dyn.* **3,** 667–669.
7. Greider, C. W. & Blackburn, E. H. (1987) *Cell* **51,** 887–898.
8. Greider, C. W. & Blackburn, E. H. (1989) *Nature (London)* **337,** 331–337.
9. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr., & Blackburn, E. H. (1987) *Cell* **51,** 899–908.
10. Cech, T. R. (1988) *Nature (London)* **332,** 777–778.
11. Panyutin, I. G., Kovalsky, O. I. & Budowsky, E. I. (1989) *Nucleic Acids Res.* **17,** 8257–8271.
12. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
13. Saenger, W. (1984) in *Principles of Nucleic Acid Structure* (Springer, New York), pp. 116–158.
14. Fowler, R. F., Bonnewell, V., Spann, M. S. & Skinner, D. M. (1985) *J. Biol. Chem.* **260,** 8964–8972.
15. Sen, D. & Gilbert, W. (1988) *Nature (London)* **334,** 364–366.
16. Arnott, S., Chandrasekaran, R. & Marttila, C. M. (1974) *Biochem. J.* **141,** 537–543.
17. Zimmerman, S. B., Cohen, G. H. & Davies, D. R. (1975) *J. Mol. Biol.* **92,** 181–192.
18. Chuprina, V. P. & Poltev, V. I. (1985) *Nucleic Acids Res.* **13,** 141–154.
19. Chou, C. H., Thomas G. J., Jr., Arnott, S. & Smith, P. J. C. (1977) *Nucleic Acids Res.* **4,** 2407–2419.
20. Donohue, J. & Trueblood, K. N. (1960) *J. Mol. Biol.* **2,** 363–371.
21. Conner, B. N., Yoon, C., Dickerson, J. L. & Dickerson, R. E. (1984) *J. Mol. Biol.* **174,** 663–695.
22. Brünger, A. T., Kuriyan, J. & Karplus, M. (1987) *Science* **235,** 458–460.
23. Blackburn, E. H. & Szostak, J. W. (1984) *Annu. Rev. Biochem.* **53,** 163–194.
24. Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyene, J., Ratliff, R. L. & Wu, J.-R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6622–6626.
25. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. (1989) *Cell* **59,** in press.
26. Sundquist, W. I. & Klug, A. (1989) *Nature (London),* in press.