APPENDIX

Alignment of Partial Denaturation Maps of Circularly Permuted DNA by Computer

IAN T. YOUNG, DONALD LEVINSTONE, MURRAY EDEN

Research Laboratory of Electronics

BIK-KWOON TYE AND DAVID BOTSTEIN

Department of Biology
Massachusetts Institute of Technology
Cambridge, Mass. 02139, U.S.A.

(*Received 18 January 1974*)

In the accompanying paper, it was necessary to align partial denaturation maps of linear but circularly permuted DNA molecules in order to determine the relative positions of the ends. As mentioned in Materials and Methods, this was accomplished by visual alignment of maps plotted on paper strips, a tedious process of possibly questionable objectivity. This Appendix describes a technique for the automated analysis by objective criteria of these maps, which yields essentially the same result with some improvement in the resolution of fine structure.

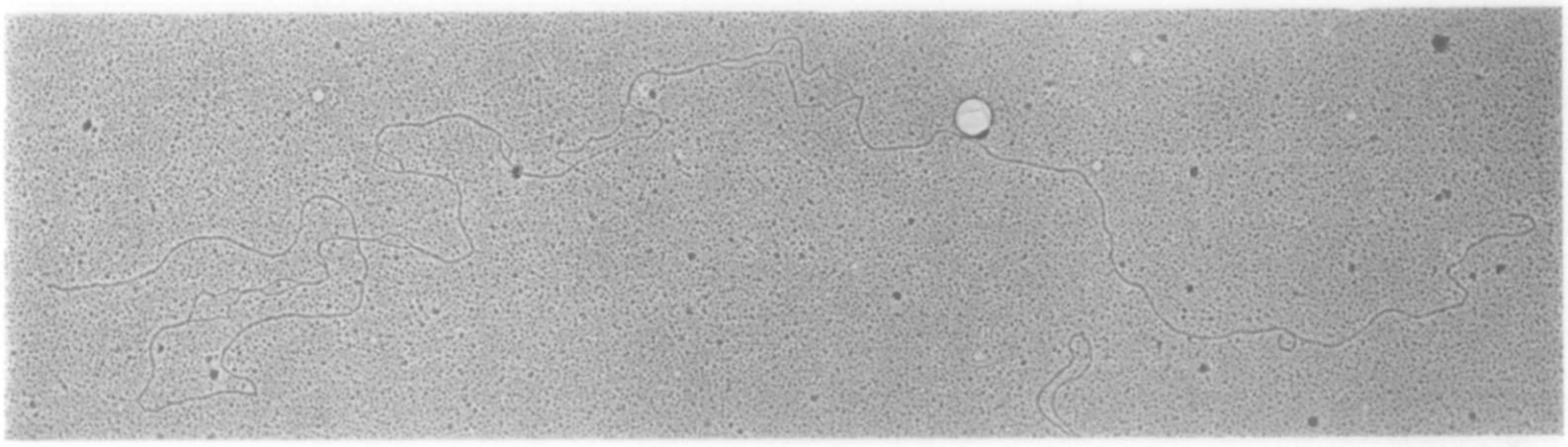
Each partially denatured P22 DNA molecule is entered into a PDP9 computer by tracing its electron micrograph on a Computek Tablet (model GT50/10). Each molecule is divided into 1024 equal segments, each of which is coded to indicate whether it is in a native (level = 0) or denatured (level = 1) region. There are 1024 such segments per DNA molecule, so that each segment corresponds to about 40 base pairs. A two-level curve $x(m)$ is thus produced, an example of which is shown in Plate AI. Of the total length, 2% is deleted from an end of each curve to eliminate the terminal repetition (Tye *et al.*, 1974). The resulting curve (which corresponds to one genome length) is treated as a circle and the displacements required to match the denaturation maps are treated as angles. The program (described below) determines the relative shifts (i.e. angles) of the denaturation maps when their numerical representations (curves) are optimally aligned.

In a set of N curves, a pairwise cross-correlation for all pairs (i,j) is performed. The cross-correlation function is given by

$$c_{ij}(k) = \sum_{m}^{L} x_i(m)\, x_j(k + m),$$

where the index m represents a running variable on the length ($L = 1024$) of the maps extended around a circle and the index k represents the relative shift (angle) between the two maps. The properties of this function have been widely studied and its appropriateness to this type of problem is well-known (Cramér, 1966; Van Trees, 1968; Lee, 1960). Wensink & Brown (1971) used a similar approach for a more limited analysis of partial denaturation maps.

(a)



(b)

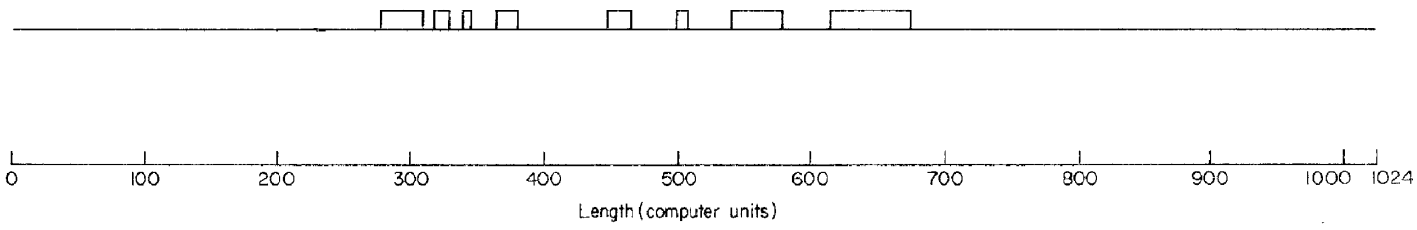
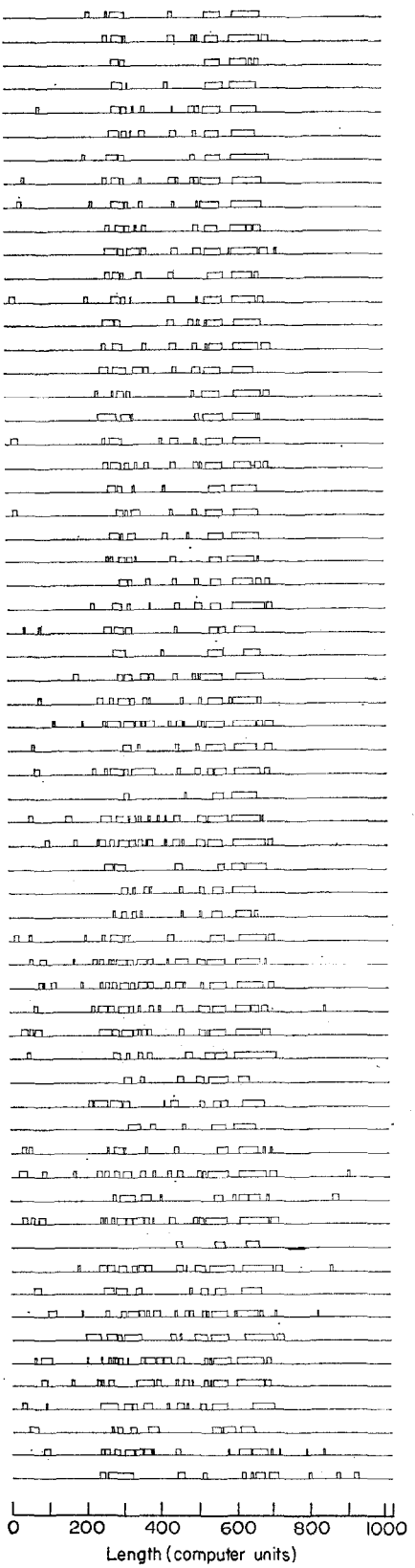


PLATE AI. (a) Photomicrograph of a partially-denatured P22 DNA molecule. (b) Computational representation of native and denatured regions.

[facing p. 528]


0
200
400
600
800
1000
Length (computer units)

produced by the computer is compared to that produced by visual alignment in Figure A2. It is obvious that the two alignments give the same pattern.

It should be noted that in the computer-generated histogram almost every molecule (52 of 63) is denatured at some point in each of the three major denatured regions, and that two of the peaks have sharp rising and falling slopes, but flat tops, indicating that essentially all (62 of 63) of the molecules have extensive denaturations at these sites. In this respect, and in terms of fine structure, the computer alignment seems more precise than the visual one.

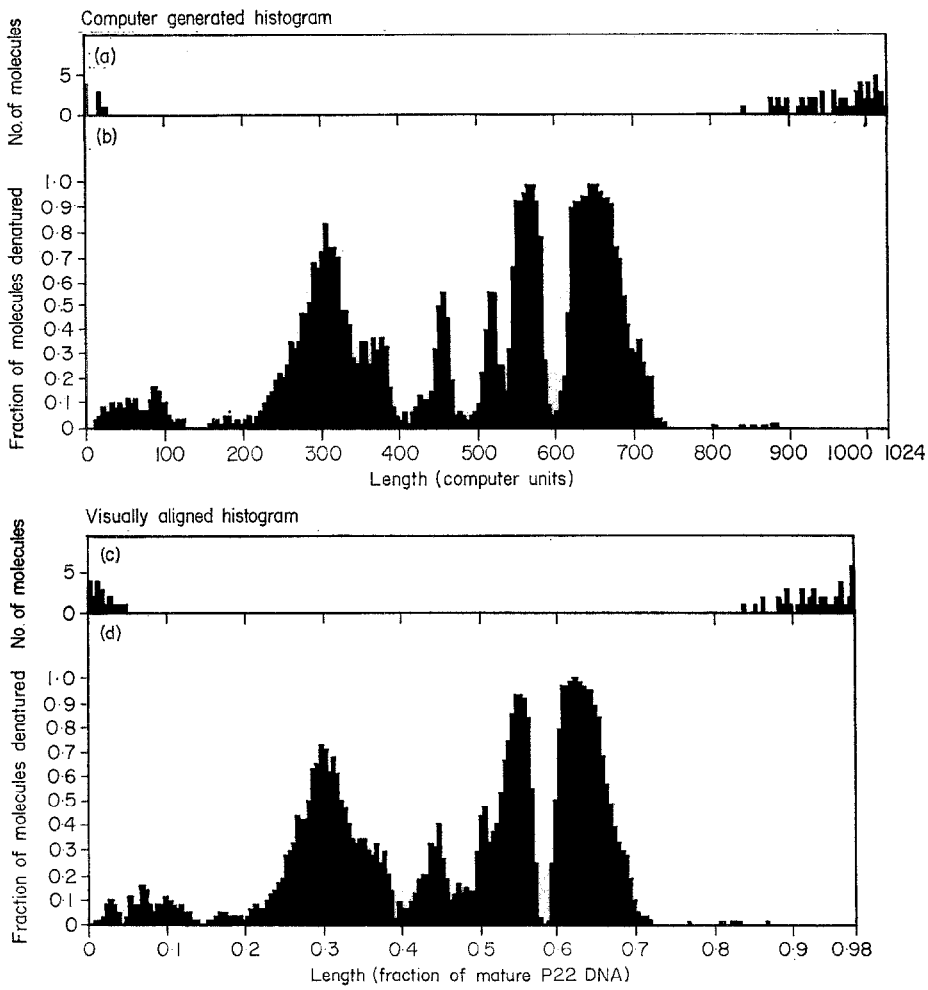


FIG. A2. Comparison of computer-generated and visual alignment of 63 partial-denaturation maps of phage P22 DNA.

(b) Histogram of the 63 curves aligned in pass 2.

(d) Original histogram of the same 63 partial denaturation maps obtained by visual alignment and permuted slightly to match the computer alignment. (a) and (c) are the distributions of molecular ends as determined from computer alignment and visual alignment, respectively.

FIG. A1. Output of second pass. Curves aligned by choice of direction and rotation parameters to maximally correlate with an average of the "best" correlated curves of the first pass.

The distribution of molecular ends in the computer-generated alignment (Fig. A2(a)) is also similar to that found by visual alignment (Fig. A2(c)); in particular, the maximum rotation to align any two curves in the final pass is 208/1024 units, or 20% of the map length.

The close similarity between the computer-generated and visual alignments of the partial denaturation maps validates both methods; the simplicity, objectivity and increased precision of alignment by the computer program recommend this approach for further work. A more detailed description of the alignment procedure including computational and statistical considerations is in preparation.

This work was supported by the National Institutes of Health under grants 5P01-GM14940-07 and 5R01GM18973-02.

We thank Eric Jensen for his assistance.

REFERENCES

Cramér, H. (1966). *Mathematical Methods of Statistics*, Princeton University Press, Princeton.

Lee, Y. W. (1960). *Statistical Theory of Communication*, John Wiley and Sons, Inc., New York.

Tye, B. K., Chan, R. K. & Botstein, D. (1974). *J. Mol. Biol.* **85**, 485–500.

Van Trees, H. L. (1968). *Detection, Estimation, and Modulation Theory*, John Wiley and Sons, Inc., New York.

Wensink, P. C. & Brown, D. D. (1971). *J. Mol. Biol.* **60**, 235–247.