A. D. HERSHEY

Idiosyncrasies of DNA structure

Nobel Lecture, December 12, 1969

In 1958, what had been learned about the genetic structure of phage particles presented a paradox. On the one hand, genetic crosses revealed only one linkage group[1]. On the other hand, physical evidence suggested that phage particles contained more than one DNA molecule and, probably, more than one species of DNA molecule. The paradox need not be dwelt on here, for it turned out that the physical evidence was mistaken: phage particles contain single DNA molecules that are species specific.

Seeking to resolve the paradox of 1958, my colleagues and I had to start at the beginning by learning how to extract, purify, and characterize DNA molecules. As it happened, our inexperience was not a severe handicap because the existing techniques were still primitive. They were primitive for good reason: until virus particles could be taken apart, nobody had ever seen a solution of uniform DNA molecules. Without knowing it we were entering one of those happy periods during which each technical advance yields new information.

Joseph Mandell and I began by attempting to make chromatography of DNA work[2]. We succeeded, as had many chromatographers before us, more by art than by theory [3]. The first application of our method, by Elizabeth Burgi and me, yielded the following results[4].

(1) DNA extracted from phage T2 proved to be chromatogmphically homogeneous.

(2) Subjected to a critical speed of stirring, the DNA went over by a single-step process to a second chromatographic species. The second species formed a single band that was not chromatographically homogeneous. We guessed that it consisted of half-length fragments produced by single breaks occurring preferentially near the centers of the original molecules.

(3) A single chromatographic fraction of the half-length fragments, when subjected to a higher critical speed of stirring, went over in a single step to a third chromatographic species that we called quarter-length fragments.

(4) Unfractionated half-length fragments subjected to stirring could be altered in a gradual but not a stepwise manner, presumably because the frag-

ments of various lengths broke at various characteristic speeds of stirring. (5) These results showed that chromatographic behavior and fragility under shear depended on molecular length, and that our starting material was uniform with respect to length by both criteria.

Burgi and I verified the above results by sedimentation analysis and took pains to isolate precise molecular halves and quarters[5].

At this point we believed we had characteristic DNA molecules in our hands but lacked any method of weighing or measuring them. Fortunately, Irwin Rubenstein and C. A. Thomas Jr. had a method of measurement but were experiencing difficulties in preparing materials. We joined forces to measure by radiographic methods the phosphorus content of T2 DNA molecules and their halves and quarters[6]. We found amolecular weight of 130 mil lion for the intact DNA. Moreover, since the DNA molecule and the phage particle contained equal amounts of phosphorus, there could be only one molecule per particle. Evidently T2 possessed a unimolecular chromosome.

With the materials derived from T2 available as standards, Burgi and I worked out conditions under which sedimentation rates in sucrose could be used as measures of molecular weight[7]. We found the useful relation

$$D_2/D_1 = (M_2/M \qquad (1)$$

in which D means distance sedimented, M means molecular weight, and the subscripts refer to two DNA species. The relation serves to measure the molecular weight of an unknown DNA from that of a known DNA when the two are sedimented in mixture. By this method the DNA of phage A, for instance, shows a molecular weight of 31 million.

Of course, equation (1) holds only for typical bihelical DNA molecules. As a check for equivalent structures, we measured fragility under hydrodynamic shear, which also depends on molecular weight and molecular structure[8].

While the rudiments of T2 DNA structure were being worked out as described above, genetic analysis of the chromosome was generating its own paradox. By this time T4 had largely replaced T2 for experimental purposes, but the two phages are so closely related that information gained from either one usually applies to both.

The paradox appeared in the work of Doermann and Boehner[9] who found, in effect, that T4 heterozygotes could replicate without segregating and were somehow polarized in structure. These properties were incompatible with the heteroduplex model for heterozygotes and eventually led Streisinger and his colleagues to postulate two radical features of T4 DNA structure: circular

permutation and terminal repetition[10]. I shall come back to these features presently, and note here only that both have been confirmed by physical analysis". I turn now to a rather different DNA, that of phage λ.

From the first, our preparations of λ DNA proved refractory in that they refused to pass through our chromatographic column and failed to yield reasonable boundaries in the ultracentrifuge. Only after we achieved sedimentation patterns of high resolution in sucrose did our results begin to make sense[12]. Then we could see in suitable preparations four distinct components. Three of these sedimented at the proper rates according to equation (1) for linear structures with length ratios 1 : 2 : 3. We provisionally called them monomers, dimers, and trimers. The fourth component sedimented faster than the monomer but slower than the dimer. We called it a closed or folded monomer.

Further analysis was possible when we found that heating to about 75° (insufficient to cause denaturation) converted everything into linear monomers. Then we could show that heating dilute solutions at 55° converted monomers entirely into the closed form. Alternatively, heating concentrated solutions at 55° yielded dimers, trimers, and larger aggregates. Moreover, the fragility of the various structures under shear decreased in the proper order: trimers, dimers, closed monomers, then linear monomers. Finally, closure and aggregation were clearly competitive processes, suggesting that each molecule possessed two cohesive sites that were responsible for both processes.

We tested our model by examining molecular halves, which by hypothesis should carry one cohesive site each, and should be able to join only in pairs (Fig. 1). This proved to be correct: by thermal treatment we could reversibly convert molecular halves into structures sedimenting at the rate of unbroken

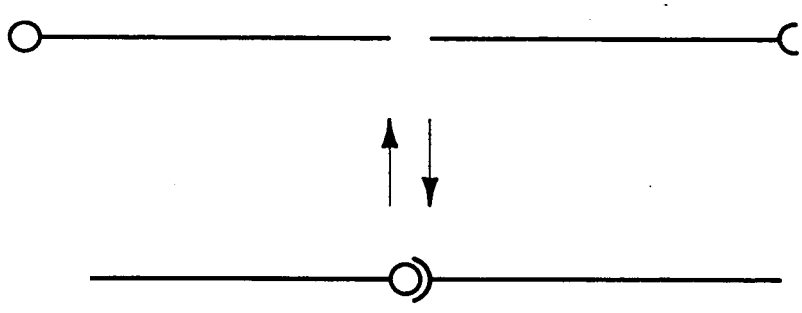

Fig. 1. Reversible joining of molecular halves of λ DNA. The upper part of the figure shows a schematic version of a single molecule cut in two. The lower part of the figure shows the halves rejoined through their terminal cohesive sites.

linear molecules. Furthermore, the rejoined halves exhibited a buoyant density appropriate to paired right and left molecular halves, not to paired right halves or paired left halves[13]. Therefore the cohesive sites were complementary in structure, not just sticky spots. Our results were also consistent with terminally situated cohesive sites, though clear proof of this came from the electron microscopists. The structures corresponding to closed and open forms of λ DNA are diagrammed in Fig. 2.

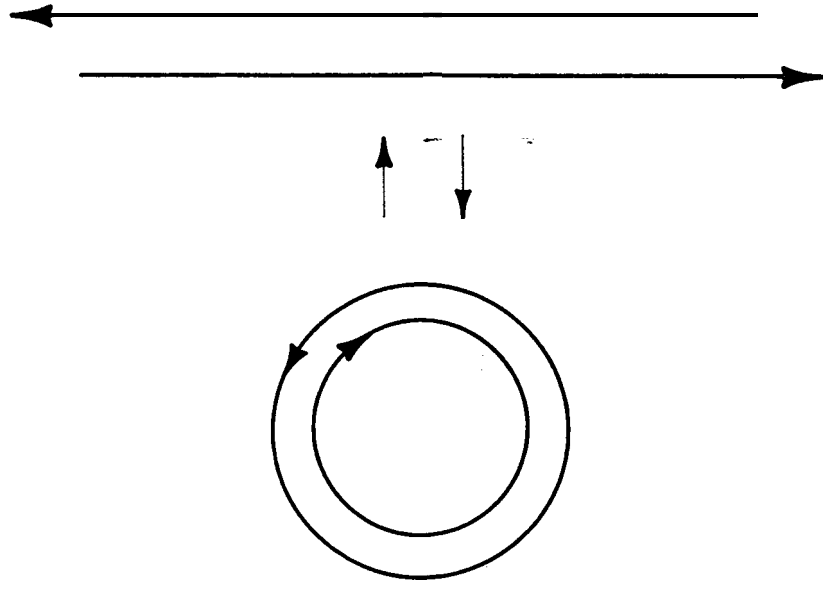

Fig. 2. Interconversion of open and closed forms of λ DNA. The bihelical DNA molecule is indicated by parallel lines of opposite polarity.

Phages T4 and λ do not exhaust the modalities of phage DNA structure[14] but they cover much of the ground. To bring together what they teach us I show in Fig. 3 idealized DNA molecules of three types.

Structure I represents the Watson- Crick double helix in its simplest form. Actually, structure I is not known to exist, perhaps because replication of the molecular ends would be mechanically precarious.

Structure II represents the linear form of λ DNA, as we have seen. Note that on paper it derives from structure I without gain or loss of nucleotides. In nature, structure II cannot derive from I because it contains the sequence za not present in I. In structure II, A and a may be called joining sequences, and they need only be long enough to permit specific base pairing. In λ, the joining

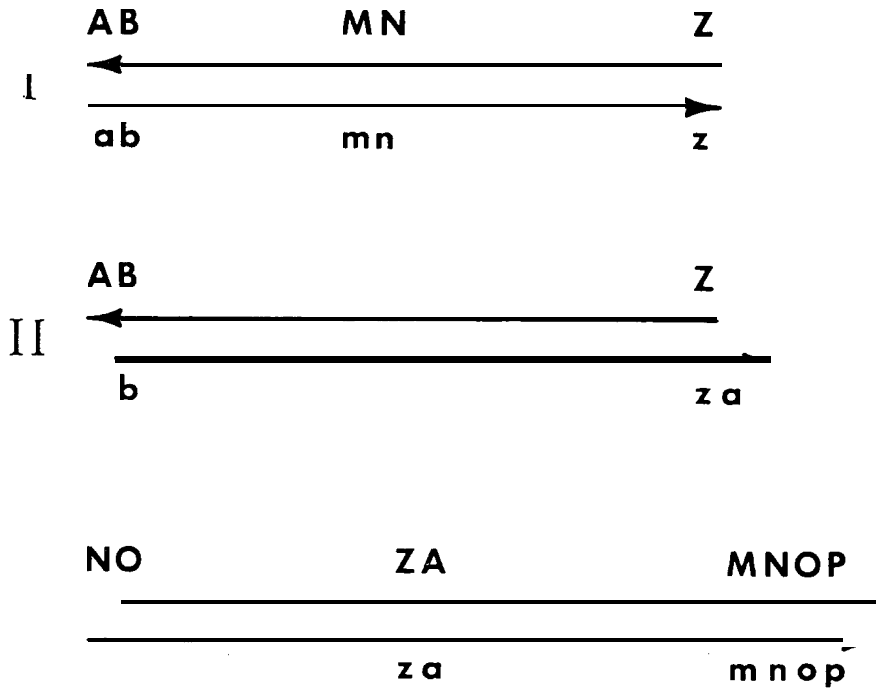

Fig. 3. Three bihelical DNA molecules represented as information diagrams. Each capital letter signifies a nucleotide sequence of arbitrary length; each small letter the corresponding complementary sequence.

sequences each contain about 20 nucleotides[15]. Presumably phage λ circumvents the difficulty of replicating ends by abolishing them (Fig. 2).

Structure III represents T4 DNA. It derives from structure I, on paper, by cyclic permutation followed by the addition of repeats at one end. T4 DNA molecules are said to be circularly permuted, meaning that the. 200000 possible cyclic permutations occur with equal frequency. The terminal repetitions in T4 DNA are relatively long: about one per cent of the molecular length containing about two genes. Thus T4 DNA solves the problem of replication of ends by making them dispensable. The same feature protects gene function during permutation, because the severed genes in a particular DNA molecule are always present in a second intact copy.

Structure III has three genetic consequences. First, molecules arising by recombination are heterozygous for markers situated in the terminal repetitions[16]. Second, such heterozygotes are likely to appear one-ended and can replicate before segregation[9]. Third, the circular permutation gives rise to a circular genetic map[10].

T4 cannot generate its permuted DNA molecules by cutting specific internucleotide bonds, and seemingly must somehow cut them to size. In principle, it could do this by measuring either total length or the lengths of the terminal repetitions. Many years ago, George Streisinger and I looked for a shortening of T4 DNA molecules as a result of genetic deletions. We didn't find any shortening. The reason is that T4 cuts its chromosome by measuring overall size. Thus a reduction of the genomic length just increases the lengths of the terminal repetitions[17].

Structure II as seen in λ DNA also has genetic consequences. Here the chromosome ends are clearly generated by cuts at specific internucleotide bonds, with the result that deletions necessarily shorten the DNA molecule[18]. The cohesive sites in λ DNA, though they do not give rise to a circular genetic map, do show up as a joint in the center of the prophage, thus giving to λ a genetic map with two cyclic permutations. In fact Campbell[19] had foreseen the need for potential circularity of the λ chromosome at a time when the DNA could only be described as goo.

Summary

The study of two phage species led to three generalizations probably valid for all viruses.

(1) Virus particles contain single molecules of nucleic acid.

(2) The molecules are species specific and, with interesting exceptions, are identical in virus particles of a single species.

(3) Different viral species contain nucleic acids that differ not only in length and nucleotide sequence but in many unexpected ways as well. I have described only two examples: λ DNA, characterized by terminal joining sequences, and T4 DNA, exhibiting circular permutation and terminal repetition of nucleotide sequences.

Epilogue

In the foregoing account I have deliberately pursued a single line of thought, neglecting both parallel developments in other laboratories and work performed in my own laboratory in which I didn't directly participate. A few of these omissions I feel obliged to repair.

Several people studied breakage of DNA by shear before we did. Davison[20] first noticed the extreme fragility of very long DNA molecules, and he and Levinthal[21] pursued the theory of breakage. However, we first noticed stepwise breakage at critical rates of shear. That observation was needed both to substantiate theory and to complete our evidence for molecular homogeneity.

Davison *et al.*[22] also showed that particles of phage T2 contain single DNA molecules, though they didn't attempt direct measurements of molecular weight.

Physical studies of DNA had of course been under way for some years before analysis of virus particles began. For instance, Doty, McGill and Rice[23] had observed a relation equivalent to our equation (1) containing the exponent 0.37. Their data covered a range of molecular weights below seven million. Larger molecules were not known at the time of their work, and could not have been studied by the existing methods anyway.

Our work on sedimentation of DNA in sucrose would have been considerably eased if we had known of the earlier work on sedimentation of enzymes by Martin and Ames[24].

The first example of a circular DNA[25], as well as the first evidence for one DNA molecule per phage particle[26], came from Sinsheimer's work with phage φX174. Its DNA comes in single strands that weigh only 1.7 million daltons, which is small enough to permit light-scattering measurements.

Elizabeth Burgi[18] demonstrated reductions of molecular weight of DNA in deletion mutants of phage λ already shown by G.Kellenberger *et al.*[27] to contain reduced amounts of DNA per phage particle.

1. A.D. Kaiser, *Virology,* 1 (1955) 424; G. Streisinger and V. Bruce, *Genetics,* 45 (1960) 1289.
2. J.D. Mandell and A.D. Hershey, *Anal. Biochem.,* 1 (1960) 66.
3. I recall a discouraging conversation with Paul Doty, in which he pointed out and I agreed that large DNA molecules could hardly be expected to reach equilibrium with the bed material during passage through a column. But at least Mandell and I were dealing with ionic interactions. A couple of years later Doty and Marmur and their colleagues were to face a real theoretical crisis: having discovered renaturation of DNA, how explain the remarkable speed of the process?
4. A.D. Hershey and E. Burgi, *J. Mol. Biol.,* 2 (1960) 143.
5. E. Burgi and A. D.Hershey, *J. Mol. Biol.,* 3 (1961) *458.*

6. I. Rubenstein, C. A.Thomas Jr. and A. D. Hershey, *Proc.Natl. Acad. Sci. (U.S.), 47* (1961) 1113.
7. E.Burgi and A.D.Hershey, *Biophys.J.,* 3 (1963) 309.
8. A. D. Hershey, E. Burgi and L. Ingraham, *Biophys.J.,* 2 (1962) 424.
9. A.H.Doermann and L.Boehner, *Virology,* 21 (1963) 551.
10. G. Streisinger, R. S. Edgar and G. Harrar Denhardt, *Proc.Natl. Acad. Sci. (U.S.),* 51 (1964) 775.
11. C. A. Thomas Jr. and I.Rubenstein, *Biopkys.J.,* 4 (1964) 94; L. A. MacHattie, D.A. Ritchie, C. A. Thomas Jr. and C. C.Richardson, *J.Mol.Biol.,* 23 (1967) 355.
12. A.D. Hershey, E. Burgi and L. Ingraham, *Proc. Natl. Acad. Sci. (U.S.),* 49 (1963) 748.
13. A.D.Hershey and E.Burgi, *Proc.Natl.Acad.Sci.(U.S.),* 53 (1965) 325.
14. C. A.Thomas Jr. and L. A.MacHattie, *Ann. Rev.Biochem.,* 36 (1967) 485.
15. R.Wu and A.D.Kaiser, *J.Mol.Biol.,* 35 (1968) 523.
16. J. Sechaud, G. Streisinger, J.Emrich, J.Newton, H.Lanford, H.Reinhold and M. Morgan Stahl, *Proc.Natl.Acad.Sci.(U.S.),* 54 (1965) 1333.
17. G. Streisinger, J.Emrich and M. Morgan Stahl, *Proc. Natl. Acad. Sci. (U.S.),* 57 (1967) 292.
18. E. Burgi, *Proc. Natl. Acad. Sci. (U.S.),* 49 (1963) 151.
19. A. Campbell, *Advan. Genet.,* 11 (1962) 101.
20. P. F. Davison, *Proc. Natl. Acad. Sci. (U.S.),* 45 (1959) 1560.
21. C.Levinthal and P.F.Davison, *J.Mol.Biol.,* 3 (1961) 674.
22. P. F.Davison, D. Freifelder, R. Hede and C. Levinthal, *Proc.Natl. Acad. Sci. (U.S.),* 47 (1961) 1123.
23. P.Doty, B.McGii and S.Rice, *Proc.Natl. Acad. Sci. (U.S.),* 44 (1958) 432.
24. R.G.Martin and B.N. Ames, *J.Biol.Chem.,* 236 (1961) 1372.
25. W. Fiers and R.L. Sinsheimer, *J.Mol.Biol.,* 5 (1962) 408.
26. R.L. Sinsheimer, *J.Mol.Biol.,* 1 (1959) *43.*
27. G.Kellenberger, M.L.Zichichi and J. Weigle, *J.MoI.BioI.,* 3 (1961) 399.