ON THE TOPOLOGY OF THE GENETIC FINE STRUCTURE

BY SEYMOUR BENZER

DEPARTMENT OF BIOLOGICAL SCIENCES, PURDUE UNIVERSITY

Communicated by M. Delbrück, September 8, 1959

From the classical researches of Morgan and his school,\(^1\) the chromosome is known as a linear arrangement of hereditary elements, the “genes.” These elements must have an internal structure of their own. At this finer level, within the “gene” the question arises again: what is the arrangement of the sub-elements? Specifically, are they linked together in a linear order analogous to the higher level of integration of the genes in the chromosome?

Until recently, the sensitivity of genetic analysis has been insufficient to answer this question. Mapping of a genetic structure is done by observing the recombination of its parts, and recombination involving parts of the structure that are very close together is a rare event. Observation of such rare events requires very many offspring and a selective trick for detecting the few individuals in which the event is recorded. It is for this reason that microorganism are the material of choice for studies of genetic fine structure, and have made it feasible to extend the fineness of genetic mapping by orders of magnitude. In favorable systems, the attainable resolution reaches the level of the molecular subunits of the hereditary material. and experimental testing of the linear arrangement of the finest structural details is therefore possible.

A number of cases have been investigated on this level.\(^2\) As a rule, closely linked mutations affecting the same characteristic can be seriated in an unambiguous way, suggesting a linear model. However, the “distances” (i.e., recombination frequencies) between mutations are not always strictly additive, and certain complexities (“negative interference” effects\(^3,\,4\)) make quantitative analysis difficult. As pointed out by Muller\(^4\) in regard to similar difficulties encountered in mapping on the chromosomal level, strict additivity of “distances” should not be taken as the criterion for the linear character of an array. A crucial examination of the question should be made from the point of view of topology, since it is a matter of how the parts of the structure are connected to each other, rather than of the distances between them. Experiments to explore the topology should ask qualitative questions (e.g., do two parts of the structure touch each other or not?) rather than quantitative ones (how far apart are they?).


\(^{4}\) Pauling, L., and R. B. Corey, these Proceedings, 37, 241 (1951).


\(^{6}\) This sample was prepared by E. R. Blout and R. Karlson. Its characterization is described in reference 8b.

\(^{7}\) An earlier estimate, based on fewer data, was 990 cal/mol. (ref. 3).

\(^{8}\) J. A. Schellman has obtained a value of 1.5 kcal. from several lines of argument (ref. 2).
In what follows, such an investigation is attempted for a small portion of the genetic structure of a virus, the "rII" region of phage T4. Using only qualitative tests, an examination is made of the topology of this structure at the molecular level.

The Material and the Method.—The methods for studying rII mutants of phage T4 have been described in detail elsewhere. Briefly, T4 phage of the "standard" form can multiply normally in either of two bacterial host strains, B or K. From the standard form of T4, rII mutants occasionally arise (detectable by their plaque morphology on B) that are defective in growth on K. An rII mutant can grow normally on K if the cell is simultaneously infected with a particle of the standard type phage. Thus, the standard type is able to perform some necessary function which the mutant cannot. Our interest is in the genetic structure that controls this particular function of the phage. The controlling structure has been traced to a small portion of the genetic map of the phage, the "rII" region, and various rII mutants can be shown to contain different blemished versions of the structure, as distinguished by certain criteria.

One criterion is the recombination test, in which two mutants are allowed to multiply within the same host cell, thereby providing an opportunity for the production of progeny which obtain parts of their genetic information from each parent. By recombination of the unblemished portions of the two mutant versions, some standard (i.e., non-mutant) individuals may be regenerated. To perform the recombination test with rII mutants, one can infect cells of strain B (in which the mutants are able to multiply) with the two mutants in question and examine the progeny for the appearance of standard type individuals by plating on cells of strain K, on which only the standard type can grow. Such recombinants will occur only if the two mutants do not contain blemishes affecting the same part of the structure. The sensitivity of the test is sufficient to resolve mutations separated by the smallest conceivable distance, i.e., that corresponding to the spacing between nucleotides in the DNA.

Thus, a simple experiment gives a qualitative yes or no answer to the question of whether or not two mutations overlap. This provides a key to the study of the topology of the structure.

The postulates upon which the procedure is based may be stated as follows:

(1) A hereditary structure in the phage determines its activity (i.e., ability to multiply in K). The standard structure consists of a set of elements, alteration of one or more of which produces a mutant (inactive) form. The subset of altered elements in any particular mutant is assumed to be continuous and not to enclose an unaltered region within its boundaries.

(2) When two or more phage particles infect the same cell, genetic recombination may occur: each element of the hereditary structure in a progeny phage particle is derived from the corresponding element in one or the other of the parental phages. The nature of the recombination mechanism, (e.g., whether involving material transfer of parental parts or a partial copying mechanism) is immaterial.

Given these postulates, the recombination test, applied to two rII mutants, reveals whether their mutations overlap or do not. By applying this test to many mutants, it should be possible to determine the manner in which the various parts or the structure are interconnected.
Effect of the topological nature of the structure: A few illustrations for model structures of different sorts may be helpful in clarifying this approach. Consider an example of an unconnected structure: a pack of cards. A complete pack of 52 subunits constitutes an active (standard) structure. The individual cards are interrelated as members of a set, but are not connected in any fixed order. A simple alteration, such as the damage or loss of one of the cards, renders the pack inactive (mutant). However, given two mutant packs, it is possible to produce a standard pack by recombination, but only if the two mutations do not intersect, i.e., the same card must not be altered in both cases. If such a test were extended to an unlimited number of mutant packs, comparing them two-by-two and in each case recording only whether a good pack could or could not be produced, the mutants would fall into fifty-two categories. Thus, the data would reveal that the standard structure consists of fifty-two distinct parts and there would be no indication of any connections between the parts.

Consider now a connected standard structure of linear topology without branches or loops: a perfect tape recording of a piece of music. Such a structure can be rendered unacceptable by a blemish—one false note, perhaps, or a blank interval (due to a jump of the tape, say). Given two independent “mutant” versions, it may be possible to fabricate a standard one by recombination, but only if the two blemishes do not overlap. The exact mechanism is immaterial—either reciprocal exchange of parts (scissors and paste) or copy-choice (partial playback) will do. Now if various mutant versions are tried, two-by-two, in each case noting only whether or not successful recombination is possible, a new sort of result may be found with this connected structure that was not possible with an unconnected pack of cards. A blemish in the recording can involve a segment of the structure (and still be “simple” according to the restriction in the first postulate). It may therefore occur that one mutation intersects two others that do not themselves intersect. Given enough defective versions, the yes-or-no results of recombination experiments would enable one to construct a linear map showing the various defects in their relative positions within the standard structure.

A linear topology places certain restrictions on the pattern of results. Representing the standard structure as a sequence of elements a b c d . . . , the structure becomes mutant upon alteration of one element or a continuous series of them.

<table>
<thead>
<tr>
<th>STANDARD</th>
<th>abcdefghijklm</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUTANT 1</td>
<td>abc</td>
</tr>
<tr>
<td>2</td>
<td>abc</td>
</tr>
<tr>
<td>3</td>
<td>dbcd</td>
</tr>
<tr>
<td>4</td>
<td>abc</td>
</tr>
<tr>
<td>5</td>
<td>abc</td>
</tr>
<tr>
<td>6</td>
<td>abc</td>
</tr>
</tbody>
</table>

Fig. 1.—(a) A linear standard structure is indicated as a series of elements. Alteration of any element or continuous group of elements produces a “mutant” structure. Six examples are given, arranged in dictionary order. (b) The recombination matrix for these six mutants. A one indicates that standard type can be produced by recombination, a zero that it cannot.
A collection of such mutants can be arranged in dictionary order according to the first altered element, as in Figure 1a. If the six mutants shown are crossed (i.e., subjected to the recombination test) in all possible pairs, the pattern shown in Figure 1b is obtained. A zero indicates that standard type cannot arise by recombination, while a one indicates that it can. Diagonal values (i.e., for crosses of mutants with themselves) are, of course, always zero, and non-diagonal values are zero or one according to whether or not the two mutations in question overlap. The matrix has a characteristic feature which is a necessary consequence of the fact that the mutants have been arranged in dictionary order. By starting from a point on the diagonal, and moving to the right (or down), one reads off the results of crossing a particular mutant with successive ones on the list. Once the first non-intersecting mutant is reached, there must be no further intersecting ones. (If there were, they would belong higher up on the list.) Therefore, the zeros in any row or column (starting from the diagonal and moving to the right or down) must form a series unbroken by ones.

Conversely, given the data obtained on a set of mutants, if the mutants can be listed in an order such that the matrix satisfies this criterion, their mutations can be represented as segments of a simple linear structure.

This is by no means possible for any set of data. A two-dimensional structure, for example, can give quite different results. In two dimensions it would be possible for four mutations to intersect as shown in Figure 2, where mutation 4 intersects with 1 and 3 without intersecting with 2. Such a situation would be impossible in a linear structure. Pairwise crosses of the four mutants would give the results shown. By listing the mutants in various ways, three different patterns may be obtained, but in no case is it possible to persuade all the zeros to form uninterrupted rows and columns.

And so on for three dimensions, branched structures, etc., each of which, if explored with a sufficient number of mutants, would give characteristic results impossible to represent in one dimension. For example, if crosses of six mutants were to give the results shown in Figure 3b, a branched structure as in Figure 3a would be required to account for them. Note that a single occurrence of a one where a zero should be is sufficient to force this conclusion. It would be an interesting mathematical problem to derive the characteristic feature of the matrix for each kind of topological space.

Experimental.—Our objective is to examine the topology of the structure in phage T4 that controls its ability to multiply in K, and specifically to make a rigorous test of the notion that the structure is linear.
The experiment consists of starting with a single particle of standard type T4 phage and isolating from it many independent, spontaneously-arising rII mutants, each of which may contain an alteration in a different part of the structure. These mutants are to be crossed with each other, noting in each case whether standard type recombinants do or do not appear. It is then to be determined what sort of structure is required to account for the results.

**Choice of non-reverting mutants:** Most rII mutants have some tendency to revert spontaneously to standard type. This is a serious limitation since, when such a mutant is used in a cross, one cannot readily distinguish standard type particles arising by recombination from those due to reversion. To circumvent this difficulty in the present experiment, only those mutants are used for which reversion has not been detected. About one in fifteen spontaneous rII mutants is extremely stable against reversion. The reversion test is most stringent, since revertants would be readily detectable on K in a proportion as low as 10⁻⁸.

Another advantage in choosing non-reverting mutants is that they are, *apriori*, most likely to include the ones in which large alterations have occurred, thereby giving a set rich in intersections.

The 145 non-reverting rII mutants used in the present study were chosen by screening some 2,000 spontaneous r mutants* of T4 and rejecting any that were not of the rII type, or were seriously "leaky" (i.e., able to grow partially on K), or reverted detectably. The chosen mutants were crossed in many pairs, in each case testing for the appearance of standard

![Fig. 3](image-url) - (a) Six possible "mutations" are shown in a branched structure. (b) A recombination matrix for these six "mutants."

![Fig. 4](image-url) - Recombination matrix for 10 rII mutants of phage T4, arranged in arbitrary order.
In an order which brings the zeros together into unbroken series. At first glance this might appear to be a formidable undertaking (the number of permutations of 19 mutants is greater than $10^{17}$), but in practice it can be readily accomplished. In Figure 5 the mutants have been successfully rearranged. Thus, all the mutations can be represented as portions of a linear space, as shown in Figure 6.

**Results:** Figure 4 gives the results for a family of nineteen mutants related to each other by intersections. These have been crossed in all possible pairs and are listed in the order in which the mutants were isolated.

Are these data consistent with a linear structure? This can be decided by attempting to rearrange the mutants in an order which brings the zeros together into unbroken series. With the procedure used (a simple spot test), the result zero for a cross indicates that, among the progeny from cells infected with the two mutants in question, standard type particles are fewer than one in about 10$^4$.
These results rule out an unconnected (pack of cards) structure, are compatible with a linear structure, and do not require any higher order of dimensionality.

The order shown in Figure 6 is not the only possible satisfactory one, since certain groups can be freely permuted without making any difference (e.g., r215, r184, rC51, r250, rC33, r782, r221, and r347). The number of uniquely ordered individuals increases as more mutants are brought into play.

In Figure 7 data are given for 145 non-reverting mutants, the mutants having been arranged in a satisfactory order, although not all the possible pairs have been crossed. All the results can be represented on a map in one dimension, as shown in Figure 8.

Functional units: The recombination test has shown that the hereditary structure controlling the ability of the phage to multiply in K consists of many parts. While alteration of any one of these leads to the same apparent physiological defect, it does not necessarily follow that all of the $r_{II}$ mutations block the same unitary function. For instance, growth in K could require a series of reactions, absence of any one of which would suffice to block the end result. It is therefore of interest to see whether the $r_{II}$ region can be subdivided in a functional sense.

It will be recalled that the needed function can be supplied to a mutant by simultaneous infection of the cell with standard type phage. Suppose that K is infected with two mutants simultaneously, as illustrated in Figure 9 (trans arrangement). If the two defects concern independent functional processes, the system should be active, since each mutant is intact with respect to the function affected in the other. However, if both mutants are defective in the same unitary function, the combination should still be inactive.

A proper control for this experiment is provided by testing the same genetic material in the cis arrangement (Fig. 9), i.e., with both mutations in one of the phages, the second phage being standard type. (In order to perform this test, the double mutant must be previously obtained by recombination.) The physiological activity of the trans arrangement is to be compared with that of the cis. If the trans is relatively inactive, a functional affiliation is indicated.

As applied to $r_{II}$ mutants, the cis test has invariably given an active system, as is to be expected when the standard type is present. The trans test divides the mutants into two clear-cut groups. Any mutant of group A can complement any one of group B, giving activity similar to that obtained in the cis configuration. Among mutants of the same group, however, the trans configuration is essentially inactive. Thus, the mutations can affect either of two distinct functional units, as defined by the cis-trans test. A functional unit so defined has been called a "cistron."

When two mutants affected in different cistrons are tested for recombination by the usual spot test, their complementary action in lysing K leads to a massive response qualitatively different from that obtained due to recombination between mutants of the same cistron. Such a massive response is indicated in Figure 7 by a two. A few $r_{II}$ mutants are defective in both functions, that is to say, complement neither A nor B group mutants, and therefore have defects affecting both cistrons.

It is important to point out that functional complementation, when it occurs, is not due to the formation of standard type recombinants inside K; full comple-
FIG. 7.—Recombination matrix for 145 rII mutants of phage T4. A two indicates a massive response, as occurs in crosses between mutants affected in different cistrons.
FIG. 8.—A genetic map deduced from the recombination data of Figure 7. Each cistron represents a separate functional unit as defined by the cis-trans test.

Correlation of recombination and function tests: How does the assignment of mutations into cistrons by the functional test correlate with their locations on the map derived from the independent recombination test? It turns out that the map may be split into two portions by the divide indicated in Figure 8 as a vertical broken line. All mutants assigned to group A by the functional test have mutations located to the left of the divide on the basis of recombination. All mutations causing functional defect B lie to the right of the divide. Those few mutants which are defective in both functions have mutations extending to both sides.
Thus, each cistron corresponds to a sharply limited segment of the linear structure.

Discussion.—In genetic mapping, mutation and recombination are both necessary, but emphasis can be placed upon one or the other. The conventional map is made by using mutations merely as “markers,” the structure being surveyed by measurements of the frequency of recombination between markers. The mapping method we have used here is a complementary one, in which the frequency of recombination is unimportant, but, in effect, mutations have been used to explore the structure.

The particular mutants used in the present work represent a special class of the spontaneously-arising rII mutants, having been chosen for their non-reverting and non-leaky character. Mutants containing large structural alterations would naturally be found in this group, and such large mutations are necessary for overlaps of the sort needed to test the topology of the structure. In fact, rII mutants that show any tendency to revert behave as though their mutations are localized to points and would have been useless in this approach. The omission of leaky mutants is perhaps part of the reason why partial complementation effects did not cause difficulty in the assignment of the mutants to clean-cut cistronic groups. This difficulty, which has occurred in some other systems, has also been encountered with some leaky rII mutants.

The postulates on which the analysis is based are, of course, idealizations when applied to an experimental system. The standard structure is not necessarily unique. There are many examples in genetics where the observable effect of a mutation may be compensated for by a “suppressor” mutation at a different loca-
tion. Thus, when two rII mutants are crossed, the double mutant formed by recombination could have the standard phenotype, thereby causing an apparent doubling of the number of standard recombinants. However, this would have no bearing on the method of mapping that has been used here, which is independent of the frequency of recombination.

One of the postulates includes the assumption that each mutation is topologically "simple." In a linear structure this would mean that a mutant should not contain two or more alterations separated by an unchanged segment. Such double mutants are, of course, possible, but were rare enough not to cause complications in this study. A stock of standard type T4 typically contains a proportion of about $2 \times 10^{-4}$ of spontaneous rII mutants. Therefore, on the assumption that each mutation occurs as an independent event, the proportion of the mutants containing a second mutation would be expected to be of the order of one in several thousand. Such double rII mutants would almost certainly not revert detectably. Among the roughly 2,000 mutants screened to select the 145 non-reverting mutants used in this study, one instance of a double mutant was in fact detected having a point defect in each of the two cistrons. When this mutant was crossed to standard type, two single mutant segregants were obtained, an A cistron one and a B cistron one. (This result does not occur for true large mutations, such as rH23.) The mutant (r928) has therefore been omitted from Figures 7 and 8. Note that the use of mutagens to raise the mutation rate is undesirable, as it would increase the proportion of double mutants.

Has the foregoing test of a one-dimensional scheme been sufficiently exhaustive? Even a set of nonsense data might happen to be compatible with a linear representation. Suppose one were to construct a matrix by tossing a coin to determine whether each "cross" gives zero or one. What is the probability that the "mutants" of such a random pattern would satisfy the criterion for a linear topology? For $n$ "mutants," the number of matrix elements to be determined is $n(n-1)/2$. Each element is either zero or one, so there exist $2^{n(n-1)/2}$ different patterns, all equally probable. Some patterns will fit the criterion for linearity, namely those in which the zeros in each row form an unbroken series, starting from the diagonal. The first row, for instance, will be satisfactory if it contains no non-diagonal zeros, or a zero adjacent to the diagonal, or two zeros adjacent to the diagonal, etc. Thus, the first row can be satisfactory in $n$ ways, the second row in $n-1$ ways, and so forth. Altogether, there are $n!$ satisfactory patterns, and the probability of obtaining one of these at the first go is $n!/2^{n(n-1)/2}$.

It may be, however, that an unsatisfactory pattern can be converted into a satisfactory one by rearrangement of the rows (and columns), and any given set of data may be subjected to $n!$ such arrangements. Although many of these will give redundant patterns, multiplying by this factor gives a maximum estimate for the chance that random data would be compatible with a linear scheme.

For $n = 19$, as for the data of Figure 4, this maximum estimate is approximately $10^{-17}$. Considering the much larger set of data for Figure 7, the possibility of a fortuitous fit with a linear topology all but vanishes.

Therefore, the observed data are manifestly not random. However, this estimate does not really apply with the same force to the question of whether the structure is linear. The problem really is: given a topological space which is other than
linear, what is the probability that this will not make itself felt (in terms of exceptions) when a given number of mutants is studied? The answer depends upon the assumptions made as to the nature of the space and the distribution of mutations within it. In the case presented above, the assumption that the matrix elements are random amounts to assuming a very complex space indeed and a distribution of mutations such that intersections are very common. This leads to a particularly low probability that the results will be compatible with a linear order.

Of particular interest is the question of branches. No case was experimentally observed of six mutants with the relationship shown in Figure 3b, which would require a branched map. However, it has been assumed that mutations have no effect upon the unaltered parts of the structure. Suppose instead that a loss along the main line which includes a branch point also necessarily leads to the loss of the whole branch. This means that in Figure 3a loss No. 1 is impossible (if the lower right arm is taken to be the branch). The remaining mutants can then be arranged in a simple line, in the order 3, 2, 5, 6, 4. More generally, if many deficiencies relating to the branch were available, one would find an apparent linear structure all along this line, with a section along the line (the portion which in reality is the branch) which is well ordered inside, but has no deletions reaching into it from the outside and stopping somewhere inside. Therefore, to eliminate the possibility that a given segment constitutes such a branch, it must be subdivided by a mutation reaching into it. This occurs for two mutants with respect to the A cistron (33 and A105), but there are none that have this property with respect to the B cistron. While this can hardly be taken as suggesting that the B cistron is a branch, it is also true that a more exhaustive study will be necessary to rule out such a possibility.\(^6\)

It is in the nature of the present analysis that the existence of complex situations cannot be disproved. However, the fact of the matter is that a simple linear model suffices to account for the data.

In confining this investigation to rII mutants of T4, attention has been focused upon a tiny bit of hereditary material constituting only a few per cent of the genetic structure of a virus and representing altogether some thousand nucleotide links in a DNA chain.\(^6\) It would seem, therefore, that the fine structure of the hereditary material, even down to its smallest molecular components, may indeed be analogous to the linear order in which the genes are integrated in the chromosome.

Summary.—The topology of the fine structure of a region of the genetic map of phage T4 is investigated by determining whether various mutations do or do not overlap. The results permit representation of the mutations as alterations in a linear structure in which the functional units defined by the cis-trans test correspond to unique segments. The possibility of branches within the structure is not necessarily excluded.

I am indebted to Mrs. Marion Sjodin for assistance in the isolation of many of the mutants, to Dr. Leslie Orgel of Cambridge University for suggesting the "dictionary order" analogy, and to Dr. Max Delbrück for suggestions regarding the possibility of branches and for his usual moderating influence. This work was assisted by grants from the National Science Foundation and the National Institutes of Health.
Demerec\(^1\) in 1946 first demonstrated that a period of time elapses between application of mutagenic treatment and the phenotypic expression of mutations induced by the treatment. Two hypotheses have developed to account for this delay. According to the first, a mutation is induced in the gene immediately on application of the mutagenic agent to that gene; and the lag occurring between application and phenotypic expression of the mutation represents that time period required for synthesis of the altered enzyme which is under the control of the mutated gene. The second hypothesis holds that induction of mutation is not immediate after application of the mutagenic agent, and that certain chemical events or cellular synthetic processes are necessary for establishment of the mutation in the genetic apparatus. According to this theory, the delay observed is the time required for these processes to take place.

It is hardly necessary to point out that these two theories are not mutually exclusive. It is readily apparent that some finite period of time will be required for synthesis of an altered enzyme following a genetic change before any manifestation of the change can become apparent. Until recently it has not been so clear that a period of time is also necessary for establishment of the mutation within the genome.

Witkin\(^2\) has cited considerable evidence that “ultraviolet induced mutations are irreversibly established only after an appreciable delay.” She has also given evidence that the induction process is complete prior to the first postirradiation cell division and indicating that the induction process is dependent on the availability of amino acids in the incubation medium during the first postirradiation hour. Haas and Doudney\(^3\) and Doudney and Haas\(^4\) have presented evidence which sug-