Seymour Benzer and T4 *r*II

Running the Map into the Ground

R Jayaraman

R Jayaraman is an Emeritus Professor of Molecular Biology at the School of Biological Sciences, Madurai Kamaraj University, Madurai, where he taught and carried out research in the area of microbial genetics for three decades.

One of the goals of genetics is to understand genes in as much detail as possible. For instance, with respect to a given gene, one would like to know its chromosomal location, its physical and genetic size, its neighbours, the number of mutations/ alleles defining the gene, the order of mutations, the genetic/ physical distance between them, etc. Thus, rather than focusing on the whole genome, one focuses on the finer details of a given genetic segment. The exercise of probing such details is called 'fine structure genetic analysis'. There are several pioneers who have contributed enormously to this area in many bacterial and phage systems. Two stalwarts, Seymour Benzer and Charles Yanofsky, stand preeminent among them. In the following pages I present briefly Benzer's outstanding work on the fine structure of the *r*II region of bacteriopage T4. These path-breaking studies contributed significantly to our understanding of the structure, organization and function of genes.

Seymour Benzer (1921–2007) is hailed as a scientist who, "more than any other single individual, enabled geneticists adapt to the molecular age" [1]. After graduating from Brooklyn College with a degree in physics, he went to Purdue University, Lafayette, Indiana, to continue his education in physics. His work at Purdue on germanium semiconductors was a forerunner to the invention of the transistor by Brattain, Bardeen and Shockley, a work which fetched them the Nobel Prize. But Benzer's mind was already set on biology. Upon the advice of the renowned microbiologist Salvador Luria, he attended the famous phage course at Cold Spring Harbor and decided that bacteriophages (bacteria-eaters, i.e., viruses that infect bacteria) would be his forte. During his post-doctoral years in the laboratory of Max Delbrück at Caltech,

in one of their famous camp-outs in the desert, Benzer met André Lwoff who invited him to the Pasteur Institute, Paris. It was in Paris that he discovered the remarkable property of the rII mutants of bacteriophage T4 and began his brilliant and path-breaking work on genetic fine structure which lasted for more than a decade. As the story goes, his research output in phage biology and genetics was so prolific that Delbrück, in a letter to Benzer's wife, asked her to tell him not to write so many papers unless there was something new to write about! This period marks his exit from molecular biology and entry into neurogenetics, a field that he founded and continued till the end. The list of honours he received is long but ironically the Nobel does not figure in it, although many believe it should have. This article focuses on Benzer's work on the fine structure mapping of the rII locus. A brief introduction to the basic principles of fine structure genetic analysis is necessary to understand and appreciate the work of Benzer.

> The ultimate fine structure of a gene is its nucleotide sequence.

## Basic Principles of Fine Structure Genetic Analysis

With the establishment of DNA as the carrier of heredity by Avery, MacLeod and McCarty, and the elucidation of its double helical structure by Watson and Crick, the abstract Mendelian 'factors' acquired a molecular identity. The gene was ultimately recognized as the sequence of nucleotides that carries the genetic information for a specific cellular function. Thus, the ultimate fine structure of a gene is its nucleotide sequence. In the absence of this information, for a geneticist, the next best thing is to have information about all the possible mutations within the gene and their interrelation. The first requirement for this exercise is a gene of choice and the availability of a large number of mutants defective in the gene, exhibiting well-defined phenotypes.

Suppose we isolate 100 mutants which are auxotrophic[1] for an amino acid X. If the auxotrophy is due to a mutation in a single gene A, all the mutants are defective in A and we will denote them as a_1, a_2, etc. On the other hand if the affected process requires two genes, A and B, some among the 100 isolates will be of the

[1] Bacteria like *E. coli* are able to synthesize most amino acids needed for making proteins (prototrophs). However, if any of the genes involved in the biosynthesis of a particular amino acid is defective, that amino acid needs to be provided in the medium for growth. Such mutants are called auxotrophs.

Figure 1. Principle of complementation analysis. If the recessive mutations M1 and M2 are located on different genes on the homologous chromosomes, the functional copies of the two genes present can give rise to wild-type phenotype (A). If the mutations M3 and M4 are located on the same gene on both chromosomes, wild-type Gene 1 function is missing and the organism will show a mutant phenotype (B).
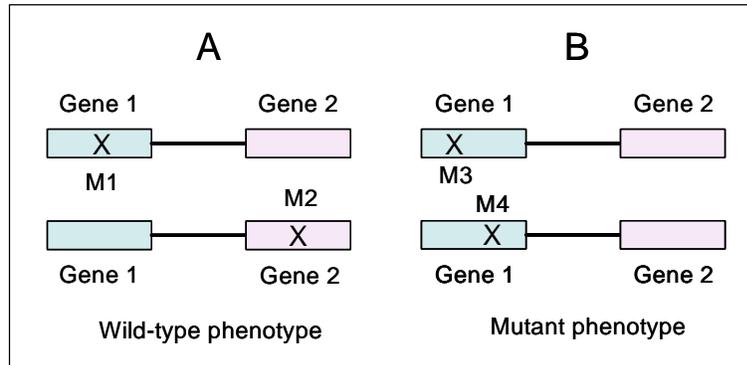
type *aB* and some will be of the type *Ab*. There could be more than two genes involved; we could have *aBC*, *AbC*, *ABc,* etc. types in the collection. How do we distinguish between these possibilities? This is done using a technique called 'complementation analysis' (*Figure* 1).

Without bothering about the technical nuances, this is done by creating a diploid cell by combining the haploid genomes of any two mutants and examining the resulting phenotype. In the case of bacteria, usually a partial diploid is created containing the mutated regions from two mutants, one on the chromosome and the other on an episome[2], making sure they do not indulge in genetic exchange (recombination). If there is only one gene involved, all the diploids will be of the type a_1/a_2 and hence will remain auxotrophic for X. On the other, hand if two genes are involved, the diploids will be of the genotype *aB/aB*, *Ab/Ab* or *Ab/aB*. It is obvious that only the *Ab/aB* cell will overcome the defect because functional copies of both the genes (*A* and *B*) are available, assuming that the mutations are recessive. This is not possible in the other diploids as they lack either the *A* function or the *B* function. Therefore all the 100 mutants could be divided into two groups called 'complementation groups'. Any two members of the same group will not complement each other but will complement any member of the other group. In other words, the number of complementation groups into which mutants defective in a given function could be grouped defines the number of genes involved in that function. For example *E. coli* mutants aux-

[2] Episomes are circular DNA molecules that can replicate independent of the chromosome. Much smaller than the chromosome, episomes can sometimes carry chromosomal genes. In such cases the cell is called a partial diploid as there are two copies of that chromosomal segment, one on the chromosome and the other on the episome.

otrophic for tryptophan fall into five complementation groups. Based on this genetic data, one can conclude that tryptophan biosynthesis needs five genes.

In the above example, let us say 50 mutants are defective in A (a_1B, a_2B, etc.) and the other 50 in B (Ab_1, Ab_2, etc.). Are all the mutants in a given complementation group mutated at the same site or at different sites? In other words, do the mutants in the same complementation group carry the same allele or different alleles? Based on their phenotypes alone it is not possible to answer this question. As before, if we construct a diploid cell in which genetic exchange is possible (recombination-proficient cell) carrying two copies of the gene bearing mutations of the same complementation group, it is possible to generate a wild-type copy of the gene by a cross-over event between the mutations as illustrated in *Figure* 2. It is not necessary to do quantitative measurements in the above experiment; a qualitative screening to see if wild-type cells arise or not with a pair of mutants will tell if the two are mutated at the same or different sites.

The major assumption is that recombination can occur within a given gene. This was not established unambiguously in the 1950s as the frequency of recombination is low for nearby loci and it needed a genetic system that has a very high resolving power. In other words, the system should be capable of recognising recombinants occurring at extremely low frequencies. Most of the genetic systems available at that time were woefully inadequate in this regard.
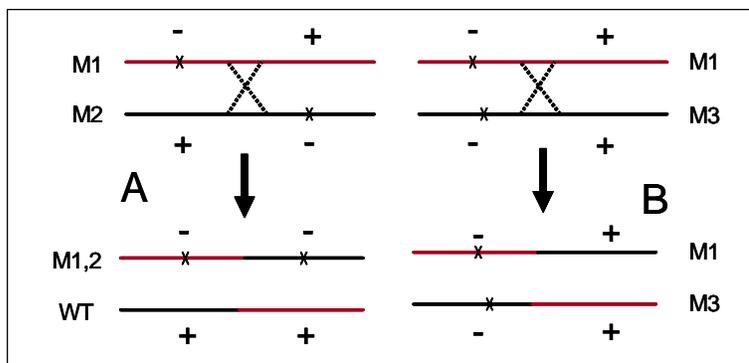
Figure 2. Generation of a wild-type gene sequence by chromosomal recombination. The dashed line represents a cross-over. The mutant sites are represented by (X) carrying the negative sign. The wild-type site is indicated by a plus sign. In A, the mutant sites on the two homologous chromosomes are different and far apart. In B, the mutant sites on the two chromosomes are the same or very close to each other and no wild type is generated. The situation is similar in bacteria, but the generation of wild type requires two cross-overs as the chromosome is circular and the incoming DNA is usually fragmentary.
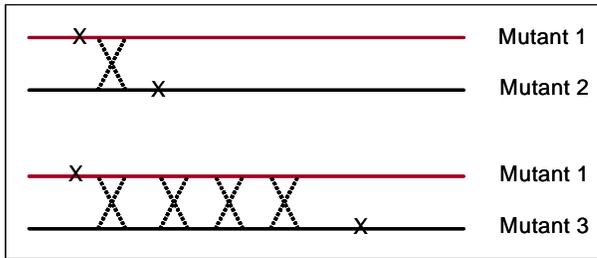
Figure 3. Recombination frequency between mutations increases with increasing distance between them as more cross-overs can occur. Thus, recombination frequency could be used to order different mutations relative to each other on the chromosome.

Suppose the 50 mutants in a given complementation group (gene) fall into ten non-identical mutated sites, each site containing a few members and the total adding to 50. For example five mutations in site 1, seven in site 2, nine in site 3, etc., all adding to 50, distributed over ten sites. The next question is: What is the spatial distribution of each site with respect to one another? This could be done by crosses (as described above) between a representative member belonging to one site with a representative member belonging to every other site and quantitatively determining the number of wild-type recombinants. The yield of wild-type recombinants in crosses between pairs of mutants is directly proportional to the distance separating them, because the probability of recombination between them increases with distance (*Figure* 3). With n sets of mutants, the number of crosses to be done is $n\,(n–1)/2$. With a set of ten mutants the number of crosses to be done is $10 \times 9/2 = 45$. If n is very large, of the order of 100 or 200, thousands of crosses have to be performed .

In the above example, crosses between mutant 1 and mutant 3 will yield more wild-type recombinants compared to a cross between mutant 1 and mutant 2. When all the exercises described above are done, we can arrange all the mutant sites in the order of increasing recombination frequency. Another important point is that recombination frequencies, like physical distances, are additive provided the distances are large, as is normally the case.

A Brief Introduction to the T4 *r*II Locus

Bactriophages or phages in short are viruses which infect bacteria. They are of two kinds with respect to their infective life style; virulent and temperate. The former, as the name suggests, always kill and lyse the infected cells. Examples of this class are the T-even (T2, T4, etc.) and T-odd phages (T1, T3, etc.). The temperate phages such as λ and P2 choose one of the two alternative modes of development, namely lytic or lysogenic modes, during
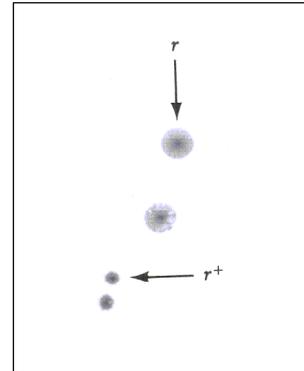
the infective cycle. In the former mode, they are similar to the virulent phages and so kill the infected cells. In the lysogenic mode, they do not kill the infected cells but integrate into the host chromosome in a benign form called the prophage. The resulting cells carrying the prophage are called lysogens. Lysogens grow normally, as non-lysogens do, carrying the prophage as if it were part of its own set of genes. However, once in a while the prophage can turn virulent. This process, called induction, occurs spontaneously in a small fraction of the population. The proportion can be increased substantially by exposure to UV and DNA damaging agents.

Phages are identified by what are called 'plaques'. When individual phage particles fall on a lawn of bacteria, they undergo several cycles of infection releasing hundreds of progeny phages during every cycle. This process causes zones of clearing on the bacterial lawn around the site where the first infection took place. They appear as holes on a dense lawn of unlysed cells (*Figure* 4).

Figure 4. Phenotypes of wild type and rII mutant of phage T4 on a lawn of E. coli B cells. (Excerpted from [5])

The *r*II Mutants of Phage T4

Wild-type phage T4 gives plaques that are 1.0-1.5 mm in diameter and have a fuzzy ring at the periphery ('halo'). This is called the wild type or r^+ morphology. Genetic variants carrying mutations in three loci called *r*I, *r*II and *r*III, show a 'rapid lysis' phenotype resulting in a different plaque morphology called the *r* type plaques. These are larger (2.0-2.5 mm in diameter) with a sharp periphery. The r^+ and *r* plaques can be distinguished just by looking at a plate containing a large number of plaques. The *r* type plaques arise at a frequency of approximately one in a million as a result of mutations in the *r* locus. Among the *r* mutants, the *r*II class of mutants have been a boon to molecular biology because of their characteristic phenotype which has proved useful in many ways. They can infect the B strain of *E. coli (E. coli* B*)* and give rise to typical *r* type plaques. But they do not form plaques on an *E. coli* K12 strain that is lysogenic for the phage λ, which is denoted as *E. coli* K12 (λ). This property is so sensitive that a single r^+ phage present in a population of 10^8 *r*II phages could be

Crossing two *r*II mutants by co-infecting the permissible *E. coli* B strain and subsequently screening the progeny on *E. coli* K12 (λ) strain, any *r*+ recombinants could be immediately detected and quantified.

detected. All that is needed is to plate a population of 10^8 *r*II phages containing one r^+ on a lawn of *E. coli* K12 (λ). The single r^+ will give a plaque! This property has been exploited by many workers in their experiments on mutagenesis, fine structure genetic analysis, genetic code, deciphering nonsense codons and their mode of action.

Benzer's Work on the Fine Structure of T4 *r*II Region

The uniqueness of *r*II mutants, namely, their inability to form plaques on *E. coli* K12 (λ) was first recognised by Seymour Benzer during 1950–52 while working in the laboratory of André Lwoff in Paris. Benzer realised that this property could be exploited to analyse the nature of genes using the *r*II mutants as a model system. He reasoned that by crossing two *r*II mutants by co-infecting the permissible *E. coli* B strain and subsequently screening the progeny on *E. coli* K12 (λ) strain, any r^+ recombinants could be immediately detected and quantified. Since very small numbers of r^+ phages in a very large population of *r*II phages could be quantitated accurately, even very small recombination frequencies could be measured. The only limitation is that if any of the *r*II mutants revert to r^+ at a high frequency, it would be difficult to distinguish between the revertants and recombinants. Therefore such mutants could not be used in crosses. In fact, many of the *r*II mutants Benzer screened for reversion turned out to be unsuitable for genetic crosses, thereby reducing the number to around 2400. But for the unique property of only r^+ but not *r*II forming plaques on *E. coli* K12 (λ), detection of r^+ recombinants would have been an almost impossible task. Benzer (1966) states: "if the phage genome were assumed to be one long thread of DNA with uniform probability of recombination per unit length, the resolving power would be sufficient to resolve two mutations even if they were located at adjacent sites". This is precisely what he accomplished, as will be discussed below.

The *r*II locus consists of two functional units, *r*IIA and *r*IIB, which Benzer called *cistrons*.

Complementation (*cis-trans*) Tests with T4 *r*II Mutants

Benzer utilised the T4 *r*II mutants in two ways. The first was to

carry out complementation tests to see how many genes define the *r*II function. For this, he co-infected *E. coli* K12 (λ) with two mutants each of which does not develop in the host and observed whether the cultures lysed or not. If they lysed, he examined the type of progeny phage (*r*II or r^+). With some pairs of mutants the infected cultures lysed and all the progeny phages were of the *r*II type, showing complementation between the mutants. With some other pairs there was no lysis and therefore, the mutants belonged to the same complementation group. This way he was able to group all the *r*II muants (over 2000) in two groups, A and B. All mutants in group A mapped at one locus and those in group B mapped at a locus very close. Therefore the *r*II locus consists of two functional units, *r*IIA and *r*IIB, which Benzer called *cistrons*. The cistron was defined as the nucleotide sequence encoding a polypeptide and is the unit of complementation. A gene may specify a single polypeptide, in which case the cicstron is the same as the gene. If a gene specifies multiple polypeptides (as in the case where multiple subunits are involved), it will consist of multiple citrons.

> The cistron was defined as the nucleotide sequence encoding a polypeptide and is the unit of complementation.

Use of *r*II Deletion Mutants to Sub-group *r*II Point Mutants

Among Benzer's collection of *r*II mutants, there were some that showed extreme stability (no reversion to r^+). Many such stable mutants also did not yield recombinants when crossed with several specific mutants. These results suggested that these mutants have suffered lesions in multiple sites (deletions) as a result of which they can neither revert nor undergo recombination with other mutants that have under-gone mutations in the same region. Benzer could determine the extent of the deletions genetically by setting up pairwise crosses with other known *r*II mutants that had already been mapped. He then exploited them to assign unmapped point mutants in *r*II to specific regions covered by the deletions (*Figure* 5).

Figure 5. Ordering of point mutations using known deletions. If the deletion D1 and the point mutation M1 within the rII locus are non-overlapping, recombination between them can give rise to a wild type phage (A). If the two mutations overlap, no recombination between them can give rise to wild-type phage (B).
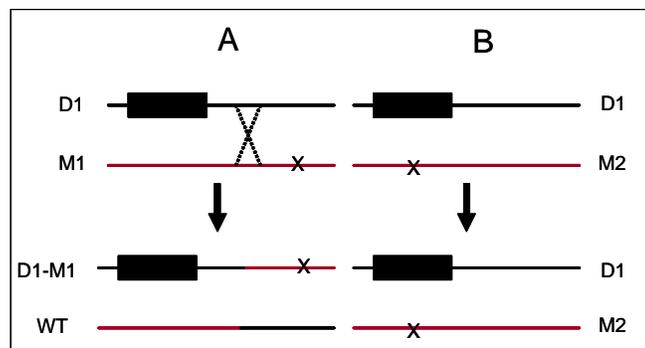
If an *r*II deletion mutant and an *r*II point mutant were crossed by co-infection in a host permissive to both (*E. coli* B), whether r^+ recombinants could be obtained or not depends upon the location of the point mutation. If it lies in the region that is missing in the deletion mutant, there is no way to generate r^+ recombinants; all the progeny phage will be only of the parental type. On the other hand if the point mutation lies in a region outside the deletion, it is possible to find r^+ recombinants among the progeny phage. Using a set of six, non-overlapping deletion mutants spanning the entire *r*II region and crossing more than 2000 *r*II point mutants with them, Benzer was able to assign each point mutant to a small sub-region. This strategy is widely used even today by geneticists working on a variety of organisms such as yeast and *Drosophila*.

Ordering the Point Mutations in a Given Sub-region

Once the approximate locations of the point mutations were determined by the method described above, the next step was to determine their relative positions within the *r*II region.

All the point mutants whose lesions had been assigned to a given sub-region of the *r*II locus were then crossed with one another to determine the (genetic) distance between them by measuring the frequency of r^+ recombinants produced. The mutations could then be arranged in increasing order of recombination frequency, giving a fine structure map of the locations of the mutant sites. This way, close to 2000 point mutations were mapped at approximately 300 sites distributed over the *r*II A and *r*II B regions. The distribution of these sites was not random. There were sites where only a few mutants mapped and also sites where hundreds of mutations mapped. In particular there were two sites, one in *r*II A and one in *r*II B, where 300–500 mutations mapped (the so-called 'hot spots'). Benzer had really "run the map into the ground".

Resolving Power of the System

The smallest recombination frequency (RF) that Benzer observed in these crosses was 0.02% (though the system had a resolving power of 0.0001%). This means that frequencies less than 0.02%

> Benzer was able to resolve two mutations located just one to two base pairs apart. This is indeed a fantastic achievement even by today's standards of high technical capabilities.

were not possible because the mutations were too close for a recombination to occur between them. Conventionally 1% RF is taken to be 1 map unit of genetic distance. Therefore the highest resolution Benzer could achieve was between two point mutations located 0.02 map units apart. Based on the principle that map distances are additive (like physical distances), the total genetic length of phage T4 has been estimated to be 1500 map units, which is the genetic size of the chromosome made up of approximately 1.5×10^5 base pairs. Therefore 0.02 map units works out to be 1.6 base pairs or 1–2 base pairs. This meant that Benzer was able to resolve two mutations located just one to two base pairs apart. This is indeed a fantastic achievement even by today's standards of high technical capabilities.

Fine Structure of T4 *r*II and Genetic Code

After the proposal of the double-helical structure of DNA, the most pressing research problem in molecular biology was the elucidation of the relationship, known as the genetic code, between the nucleotide sequence of the gene and the amino acid sequence of the protein it codes for. Although Benzer was not directly working on the 'cracking' of the genetic code, as Crick and others were, his work had a great impact on the same. The characteristic phenotype of T4 *r*II mutants which Benzer discovered was brilliantly exploited by Crick to elucidate the triplet nature of the genetic code. Crick also used the *r*1589 mutant (see below) which Benzer had isolated (in addition to his own isolates of *r*II mutants) to substantiate the nature of the genetic code.

Benzer and his student Champe had isolated a very interesting *r*II mutant, namely, *r*1589 in which a bit of the right end of the *r*II A cistron and a bit of the left end of the *r*II B cistron were deleted. This resulted in the fusion of the two cistrons and loss of *r*II A function; however the *r*II B function was not affected. The *r*1589 mutant could complement *r*II B mutants in mixed infection but could not complement *r*II A mutants. Benzer and Champe (as well as Crick and co-workers) found that some point mutations in the A part of the *r*1589 mutant abolished the B complementing

Benzer had really "run the map into the ground".

Suggested Reading

[1] W Hayes, *Genetics of bacteria and their viruses*, **2nd Edition, John Wiley & Sons, New York, 1968.**

[2] T D Brock, *The emergence of bacterial genetics*, **Cold Spring Harbor Press, Cold Spring Harbor, New York, USA, 1990.**

[3] S Benzer, *In phage and the origins of molecular biology*, **pp. 157-165. Cold Spring Harbor Press, Cold Spring Harbor, New York, USA, 1966. (This is a beautiful account, in Benzer's own words, on how he got to work on the T4 *r*II system).**

[4] W A Harris, *Seymour Benzer (1921-2007): The Man who took us from Genes to Behaviour. PloS Biol.* **6(2). e41 (A brief obituary essay on the Life, Work and Achievements of Seymour Benzer).**

[5] G S Stent and R Calendar, *Molecular Genetics: an Introductory Narrative*, **W H Freeman & Company, San Francisco, USA, 1978.**

[6] Seymour Benzer, The Fine Structure of the Gene, *Scientific American*, **Vol.206, No.1, pp.70–80, 1962.**

Address for Correspondence
R Jayaraman
School of Biological Sciences
Madurai Kamaraj University
Madurai 625 021, India.
Email:
aarjay007@rediffmail.com

activity while some others did not. They suggested that the former class of mutations are of the 'nonsense' type, resulting in the premature termination of translation as a result of the introduction of a stop codon and the latter are of the 'missense' type where one amino acid is substituted by another. Both groups made another remarkable observation that those mutants that showed a nonsense phenotype in some hosts were able to complement *r*II B mutants in some other hosts. That is, nonsense codons could be read as sense codons by some strains. They are, therefore, suppressible. It is remarkable that these conclusions, arrived at long before the genetic code was deciphered, were experimentally verified later when more facts about the genetic code became known. As Harris (2008) has succinctly stated, Benzer "discovered so much and, in so doing, opened up so much to discover".

That is enough Seymour. Stop that and do something more worthwhile