Essay

What has phage lambda ever done for us?

Noreen E. Murray[1] and Alexander Gann[2]

Shortly after the dawn of biochemical genetics, *Escherichia coli* K-12 replaced *Neurospora crassa* as the key model organism. With *E. coli* K-12 came another, even simpler, system: each bacterial cell contained a dormant virus (bacteriophage) called lambda (λ). Occasionally, the quiescent λ genome was activated to generate free phage particles. Thus it was discovered that phage λ propagates by two alternative pathways: lytic or lysogenic. In lysogeny, it is now known that the phage genome is integrated within the bacterial host genome and replicates passively along with it; in the lytic pathway, the λ genome replicates free from the bacterial genome, directs the production of phage capsid proteins, and kills its host to release some 100 or so new phage particles (Figure 1).

The fortuitous discovery of λ early in the development of molecular biology resulted in this small virus becoming a model system for studying the molecular basis of fundamental biological processes. In this essay, we discuss how the genetic and biochemical analyses of the developmental pathways of λ proved so influential (see also Box 1). In particular, we focus on the control of gene expression; the influence of phage morphogenesis on our understanding of protein folding; and DNA recombination by homologous, illegitimate and site-specific mechanisms. Studies with λ have contributed much to our understanding of the molecular basis of these processes, their biological and evolutionary roles, and how they have been harnessed by experimenters, most particularly in the development of recombinant DNA technologies.

Gene regulation

Phage λ was one of the two model systems whose study revealed the basic concepts and mechanistic details of gene regulation. These two systems — the other was the *lac* operon of *E. coli* — were studied in parallel at the Pasteur Institute in Paris in the 1950s and early 1960s. Each was of biological interest in its own right, and there seemed at first no reason to believe that the way the presence of lactose triggers *E. coli* to take on the ability to digest that sugar would be mechanistically related to how a λ phage 'chooses' between lytic and lysogenic growth upon infection of an *E. coli* cell.

The two problems were dissected using similar genetic approaches — in particular, by isolating mutants that were constitutively active for a given behaviour in each system. Thus, mutant *E. coli* strains were isolated that produced β-galactosidase in the absence of lactose, and mutant phage that grew lytically even under circumstances when they should grow lysogenically. For both *lac* and λ, the mutants fell into two distinct classes: one class defined a *trans*-acting factor, called the repressor; the other, a site on DNA, the operator. Analyses of these mutants in a variety of elegant experiments led to the interpretation that, in both systems, behaviour was controlled by regulation of gene expression, and, specifically, that each system uses its repressor to switch off expression of relevant genes, acting through the operator sites near those genes. Under suitable conditions, in the presence of lactose in the case of the *lac* genes for example, the repressor is inactivated, and the genes expressed.

It was not at the time clear whether the repressors were RNA or proteins. They turned out to be proteins, the first firm evidence being amber nonsense mutations in the λ repressor gene. Both repressors were subsequently isolated and shown to bind directly to their operator sites on DNA.

DNA recognition and cooperative binding

Working out how proteins can recognize specific DNA sequences was critical to understanding many processes, including gene regulation. Even in higher eukaryotes, where there are additional complexities

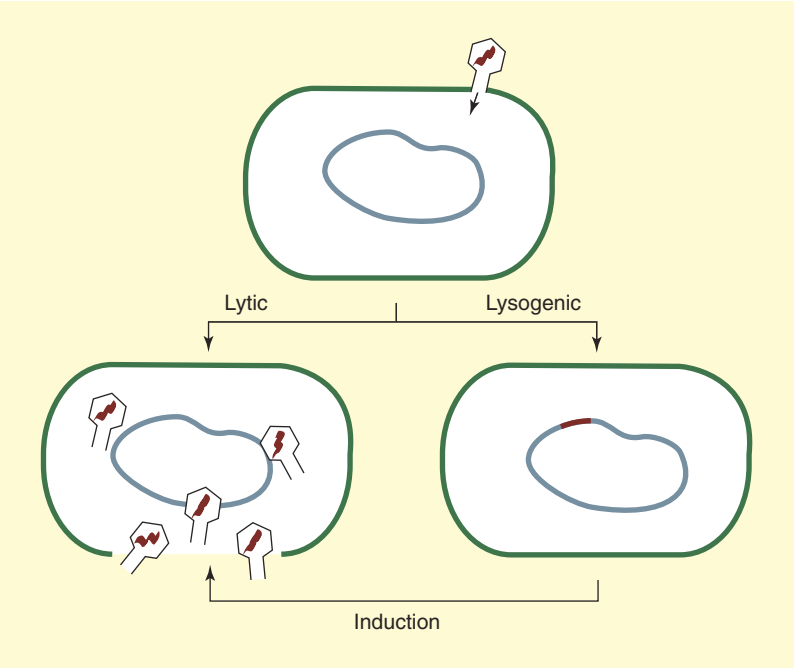

Figure 1. The life cycle of λ.

Upon infection, the phage can pursue lytic growth or lysogeny. A lysogen is stable, unless it receives a signal to induce, when essentially all lysogens switch to lytic propagation. The bacterial chromosome is in blue, that of λ in red.

Box 1

Other contributions of λ.

Phage λ has made many contributions beyond those highlighted in the text. *E. coli* lysogenic for λ provided the selection system for the first refined fine-structure analysis of genes — in this case the *rII* genes of phage T4. This system then enabled identification of the triplet nature of the genetic code.

The rapid progress in many aspects of biology during the past 25 years owes much to recombinant DNA technology, and hence to the discovery of restriction enzymes. These were first recognised as a barrier to the transmission of phages λ and P2 — it was this phenomenon that gave rise to the name 'restriction'. Experiments with λ indicated that the 'barrier' was an enzyme that attacked incoming DNA in the absence of a strain-specific modification. λ DNA provided the substrate for the purification of the first restriction endonuclease.

The size and nature of the λ genome, and ease of purification, made it a common substrate for detecting, purifying and analysing the enzymes, such as polymerases, endonucleases and ligases, that underpin recombinant DNA. In the early 1970s, SV40 and λ were used in the development of agarose and polyacrylamide gel systems for the separation and visualization of DNA fragments. Phage λ DNA was chosen to demonstrate the feasibility of 'shotgun' sequencing of a genome, an achievement published in 1982. λ is a commonly used single-molecule substrate for analysis of protein-DNA interactions, particularly those involving translocation, and provides a system for estimating mutation frequency in transgenic rodents.

associated with chromatin and DNA modifications, the action of sequence-specific DNA-binding proteins lies at the heart of gene regulation. Work on λ repressor was critical in uncovering how a protein recognizes a specific DNA sequence. In addition to sequence recognition, there is another feature typical of DNA-binding proteins: often two or more proteins binding to nearby sites interact with each other, and so bind to those sites cooperatively. This, as we shall see, heightens affinity and specificity of binding, and can be used to ensure steep 'all-or-none' responses. It was again studies of λ repressor that revealed the existence, and mechanism, of cooperative binding.

The crystal structures of λ repressor and Cro (another λ-encoded repressor) were determined in the early 1980s. A region of secondary structure predicted to recognize DNA was identified in each repressor, and was named a helix-turn-helix motif. It was suggested that this motif recognizes DNA sequences by inserting one of the helices — the so-called recognition helix — into the major groove of DNA. Amino-acid side chains protruding from the recognition helix make specific contacts with the edges of base pairs, thereby distinguishing between specific sequences.

Confirmation that these repressor proteins really did recognize DNA in this manner came initially from genetic studies, including one in which the specificity of a repressor from a phage closely related to λ (phage 434) was switched to that of a third lambdoid phage (P22) by making the recognition helix of 434 resemble that of P22. Later, direct demonstration of DNA recognition by the recognition helix came from the X-ray structures of the phage repressors bound to oligonucleotides containing their respective operators.

We now know, from the structures of many other DNA-binding proteins, that the helix-turn-helix motif is a common means of DNA recognition. This mechanism extends beyond bacteria, and includes, for example, a closely related motif found in the homeobox proteins of organisms as diverse as yeast, flies and mammals. More generally, even when the exact domain used to recognize DNA is not a helix-turn-helix, it nevertheless often involves recognition by insertion of an alpha helix into the major groove of the DNA — for example, in the case of zinc finger, leucine zipper and helix-loop-helix domains.

DNA-binding proteins generally bind as dimers, recognizing rotationally symmetrical DNA sequences. Dimerization ensures that both the specificity with which the correct site is recognized and the affinity with which it is bound are high. In a further extension of this, proteins often bind DNA cooperatively. Thus, λ repressor recognizes a typical site as a dimer, but it can bind adjacent sites as a tetramer — that is, a dimer binds to each of the two sites, but the two dimers interact with each other as they do so, further increasing specificity and affinity.

To illustrate this, consider O_{R1} and O_{R2} — two operator sites for λ repressor in the phage genome. When these two sites are isolated from each other and examined individually, repressor binds O_{R2} only when present at a ten-fold higher concentration than that needed to bind O_{R1}. But when the sites are adjacent on the same DNA molecules as they naturally are in the genome, repressor binds both simultaneously, and at the lower concentration. Furthermore, it has recently been shown that, in a lysogen, λ repressor can even bind as an octamer to four sites simultaneously. Two of these sites are separated from the other two by over 3 kb, and binding of the octamer of repressor to all four sites requires the DNA between the separated pairs of sites to form a large loop.

How does cooperative binding increase affinity? Consider O_{R1} and O_{R2} (Figure 2A). Repressor bound at O_{R1} simultaneously interacts with repressor at O_{R2}. Periodically, repressor will dissociate from O_{R2}, but be held in the vicinity of that site by its continued interaction with repressor at O_{R1}. Thus, it is held at a high local concentration and will rebind the site efficiently. Only on the rare occasion when repressor lets go of O_{R1} and O_{R2} at the same time — or repressor at O_{R1} lets go of the DNA and of repressor at O_{R2} simultaneously — will repressor dissociate and be free to drift far from the site.

How does cooperative binding increase specificity? For a pair of proteins to bind cooperatively they must be bound to sites relatively close to one another. Repressor will periodically, and fleetingly, bind to incorrect sites — but in the context of a whole genome, rarely will two non-specifically bound molecules be near enough each other to bind cooperatively.

Repression and activation

How do transcriptional repressors and activators work? Once again, λ provided an archetypal picture. λ repressor bound to a site that overlaps a promoter switches off expression by excluding binding of RNA polymerase. Many bacterial repressors work in this way, though not all. λ repressor is an activator as well as a repressor, and uncovering how it works as an activator revealed a mechanism used widely in bacteria and eukaryotes.

By binding to a site adjacent to a promoter, and making a direct contact with RNA polymerase, an activator can recruit that polymerase into a stable complex at the promoter, and this suffices to stimulate transcription. This is another example of cooperative binding. Several experiments showed that λ repressor works this way to activate its own promoter, P_{RM} (Figure 2A). Perhaps the most significant of these experiments was the isolation of mutations that defined a region on the surface of repressor — its so-called activating region — which makes contact with polymerase. These mutations did not alter DNA binding by λ repressor, but eliminated its ability to activate. Comparable mutations that altered the surface of RNA polymerase touched by the activating region were also described — these eliminated activation by wild-type λ repressor.

In a further experiment, it was shown that recruitment of polymerase to the P_{RM} promoter through a protein–protein interaction is all that is required for activation. This was done by taking two proteins known to interact with each other, and attaching one to a DNA-binding protein and the other to polymerase. This engineered polymerase can be activated by the makeshift activator, even though the protein–protein interaction used was one not normally involved in activation. Genes in many organisms can be activated by this simple mechanism of recruitment.

Finally, in the case of λ repressor, the crystal structure of repressor bound to DNA, and interacting with its target region on RNA polymerase, shows a direct protein–protein interaction that does not cause a conformational change in either partner.

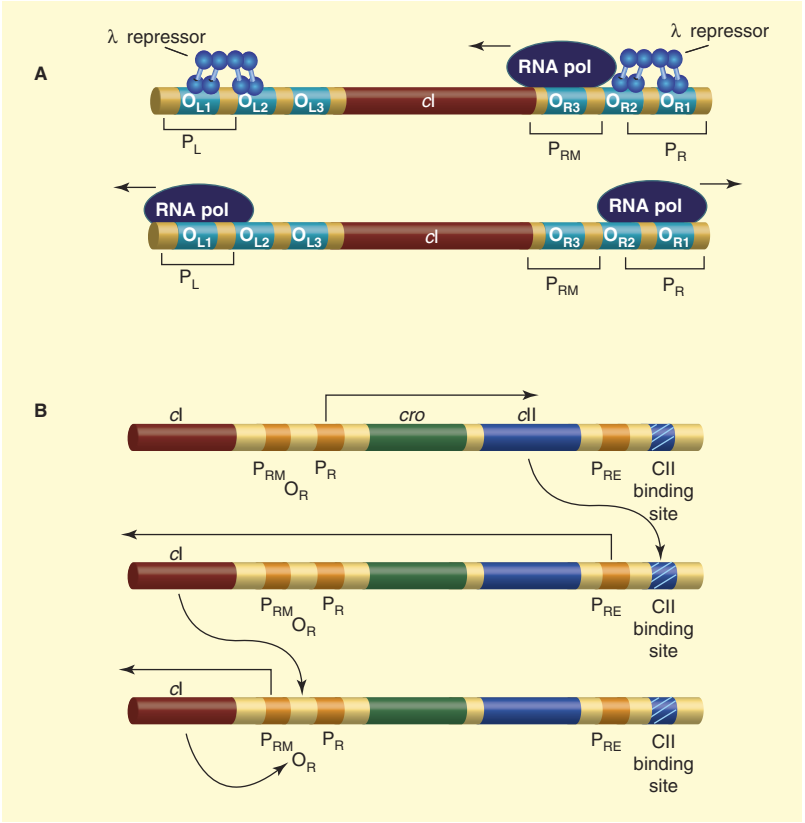

Figure 2. Gene regulation in λ.

(A) The binding of repressor and RNA polymerase to various sites within the left and right operator regions. RNA polymerase is shown bound at the three promoters in this region. The top line shows the arrangement found in a lysogen, the bottom line shows polymerase transcribing from P_R and P_L upon initial infection. In the lysogen, the repressors bound at O_R and those at O_L would interact to form a DNA loop, but for clarity this is not shown here. (B) The major component of the regulatory network required for a phage to establish lysogeny (see text). The figure ignores the role of *cro* (shown in green). This gene encodes a second repressor which binds to the same sites as λ repressor, but with different affinities and consequences. Transcription from P_R leads to the production of CII, an activator essential for lysogeny, and Cro, which opposes that pathway (by binding to O_{R3} and O_{L3}). The resolution of this competition is important in determining whether the phage develops lytically or establishes lysogeny. Other features of the decision-making apparatus have been widely studied, but are not discussed here, including the following. Transcription from P_L produces CIII, a protein that protects CII from degradation and thus promotes lysogeny. CII, in addition to activating transcription of *cI*, also activates transcription from two other promoters, one directing expression of a gene required for integrating the phage genome into the bacterial chromosome, the other an antisense RNA that opposes expression of gene Q whose product is required for expression of the late (lytic) genes. The Q protein works as an antiterminator of transcription; a second phage-encoded antiterminator (N) is important in regulating early events in infection. The binding sites, promoters and genes are not to scale.

Lambda as systems and synthetic biology

In addition to teaching us about basic mechanisms of gene regulation, λ also provides an unusually complete picture of a gene regulatory network — its description is, in a sense, the original piece of systems biology. Thus, λ has been a focus of much mathematical modeling, and has offered an ideal system in which to develop and test such approaches.

More recently, a working version of the λ switch has been created using heterologous components. In this example of synthetic biology, λ repressor is replaced by the bacterial Tet repressor, Cro by Lac repressor, and the various λ operator sites are modified to be recognized by one or other of these new repressors. This construct demonstrates the power of such an approach — showing it is possible to recreate complicated behaviours using simple, well understood, components.

What follows is a brief description of how part of the natural λ system works. Upon infection, λ may pursue either lytic or lysogenic growth (Figure 1). The early stages of infection are the same whichever pathway will eventually be followed. Thus, as soon as the phage genome enters the cell, the two strong promoters P_L and P_R initiate high levels of expression (Figure 2A, bottom). These two promoters direct the expression of many genes needed for lytic development, but also a key regulator of lysogenic development — the product of the *cII* gene. CII is a transcriptional activator that binds adjacent to, and activates transcription from, the promoter P_{RE}. The transcript includes the *cI* gene, and so CII stimulates synthesis of λ repressor (Figure 2B).

For lysogeny to be established, repressor must reach a level high enough to switch off transcription from P_R and P_L, and maintain expression of its own gene (Figure 2A, top). Repressor can bind to six sites arranged in two operator regions, O_R and O_L. At low concentrations, repressor binds O_{R1} and O_{R2} (cooperatively, as we saw), and similarly O_{L1} and O_{L2}. In this arrangement, transcription from P_R and P_L is switched off, while further expression of repressor is stimulated from the promoter, P_{RM}. This expression of *cI*, activated by repressor itself bound at O_{R2}, is an example of auto-positive regulation. When repressor levels get too high, it also binds to O_{R3}, switching off transcription from P_{RM} until levels drop again — an example of auto-negative regulation. The dual regulation of repressor synthesis — auto-positive and auto-negative — ensures that the lysogenic state, once established, is stable.

There are several additional layers to this decision-making process, some of which are touched on in the legend to Figure 2. Many of the mechanistic and strategic features we learnt from λ are seen in other systems, including, for example, *Drosophila* development. There, one again finds networks of genes regulated by site-specific DNA-binding proteins, cooperative binding, auto positive and negative regulation, and alternative promoters for the establishment and maintenance of gene expression patterns.

The epigenetic switch

Once a lysogen is established it is stable, with the phage genome, in this state called the 'prophage', remaining dormant almost indefinitely. But, if the cell is threatened, this state can be reversed in essentially every cell, and the prophage switches into the lytic cycle by a process called induction (Figure 1).

Maintenance of the lysogenic state is an example of epigenetic gene regulation. By epigenetic we mean that a pattern of gene expression is maintained through multiple generations in the absence of the signal that initiated it, and without change to the DNA sequence. There is much talk of epigenetics these days, particularly in the field of eukaryotic gene expression, and it is often assumed that the mechanism must involve chromatin or DNA modification. But λ provides a mechanistically transparent example that clearly requires neither.

What are the features of the lysogenic state that ensure both its remarkable stability and, at the same time, its acute sensitivity to induction? We have already described the tight regulation of repressor levels in the lysogen (through cooperative binding, and auto positive and negative regulation). Repressor concentration must be kept high enough to keep the lytic genes turned off, but not so high that induction is impeded.

How does induction work? Treatments such as UV irradiation cause DNA damage and threaten the cell. This activates a cellular protein called RecA which stimulates proteolytic cleavage of λ repressor, separating the part that mediates cooperative interactions, including dimerization, from the part that mediates DNA binding. This eliminates cooperativity and causes rapid loss of repressor binding to O_R and O_L. In consequence, P_{RM} is no longer activated and transcription from P_R and P_L is permitted (Figure 2A). The prophage is then excised from the host genome and enters the lytic cycle. Use of proteolytic cleavage of regulators to trigger rapid cellular responses is widespread, and examples include apoptosis in animal cells.

Chaperones and protein folding

The correct folding of polypeptide chains is fundamental to the biological role of proteins. The mechanism by which a polypeptide folds was of major concern long before misfolded proteins became relevant to understanding some degenerative diseases. Initially, the influential principle of 'self-assembly' was derived from experiments in which it was demonstrated that a denatured protein could refold spontaneously *in vitro* to regenerate the correct three-dimensional structure. However, experiments examining Rubisco, an oligomeric enzyme in chloroplasts, led to the suggestion that cellular proteins, described as 'molecular chaperones', have a critical role in ensuring the folding of some polypeptide chains, and the subsequent assembly of the folded chains into oligomeric structures. Confidence in this model was catalysed by elucidation of a bacterial function necessary for phage morphogenesis.

In the 1970s, mutant strains of *E. coli* were reported which prevent the growth of wild-type phages by interfering with their morphogenesis. But one group of such mutants could support the growth of derivatives of λ with particular changes in the E protein, the major component of the phage capsid. These bacterial strains were defined as having a GroE (for 'Grow E') phenotype. Cloning of the relevant coding region identified two genes, *groEL* and *groES*.

When the predicted amino-acid sequences of GroEL and the large subunit of the Rubisco-binding protein were compared they were found to be 50% identical. GroE was already implicated in protein assembly of the phage capsid. Furthermore, GroEL, like the Rubisco-binding protein, is oligomeric: it was known to comprise 14 subunits. The sequence similarity of Rubisco-binding protein and GroE, enhanced by what was known about the structure and role of GroE, provided striking support for the early novel ideas about chaperones.

The combination of *in vivo* and *in vitro* experiments, including determination of the three-dimensional structures of the GroEL complex, have done much to elucidate the mechanisms of a molecular chaperone. The GroEL oligomer forms a cage lined with hydrophobic residues. GroES binds to GroEL in an ATP-dependent reaction. This binding has two effects: it triggers large conformational movements in GroEL which enlarge the cavity to a size that can accommodate polypeptides up to 60 kDa, and it results in GroES sitting on top of the cavity. Within this cage, a polypeptide can complete its folding while being protected from other polypeptide chains.

From these studies emerged the idea that many proteins do not, as had long been assumed, fold spontaneously in cells, but rather they require the assistance of chaperones.

The molecular and genetic foundations of homologous recombination

The susceptibility of phage λ to both genetic and physical analyses has been critical to our understanding of mechanisms that recombine DNA. Phage λ provided an effective substrate for assaying recombination frequencies within segments of a relatively short linear genome. But its genome was also amenable to density labelling and equilibrium centrifugation, a technique that permitted quantification of the parental DNA acquired by recombinant phages. Pioneering experiments in 1961 initiated this technology to address the topical question of whether recombination occurred by the mechanism of 'copy-choice' — by DNA replication switching from one parental template to another. The experiments demonstrated that one class of recombinants, from a cross between 'heavy' and unlabelled phages, had acquired 86% of their DNA from the 'heavy' parent, a finding consistent with the generation of recombinants following the breakage of parental molecules. Recombination could not, therefore, be the exclusive result of copy-choice replication.

In the 1960s, geneticists identified genes essential for homology-dependent recombination. The *recA*, *recB* and *recC* genes of *E. coli* identified some contributors to a bacterial pathway, a key component of which is now known to be the RecBCD complex. A simpler pathway for λ was identified by mutations in the *redα* and *redβ* genes. Recombination by any pathway requires, as its final step, covalent joining of the DNA strands. DNA ligase, the relevant essential enzyme in *E. coli*, was identified in 1967 using the cohesive ends of λ DNA as a substrate. In the same year, the *redα* gene of λ was shown to encode an exonuclease that acts at the ends of double-stranded DNA, degrading in a 5′ to 3′ direction to produce single-stranded termini. In 1971 the purified products of the *red* genes of λ, followed by DNA ligase, were used to generate recombinant λ genomes *in vitro*. These experiments demonstrated the generation of heteroduplex joints by the pairing of complementary single-stranded DNA, and the cessation of exonuclease activity when the 3′ terminus has been assimilated to create a nick. DNA ligase then seals the nick to create continuity. Thus, λ exonuclease prepares the ends of DNA molecules for recombination; DNA ligase completes the process. The product of *redβ* was subsequently shown to be essential for the annealing of homologous single-stranded DNA.

Studies of λ revealed links between recombination and replication. In 1966, it was shown that, although λ DNA initially circularizes on infection and replicates as a circle, late replication yields linear concatemeric molecules. The rolling-circle model of DNA replication explained this observation. In the early 1970s, experiments monitoring replication of the λ genome demonstrated that loss of any of three λ genes — *redα*, *redβ* or *gam* — impaired the production of concatemeric DNA.

The role of Gam was explained by its inhibition of exonuclease activity. DNA replication of the circular genome, on encountering a nick, could lead to a break in the λ genome, the initiation of rolling-circle replication, and susceptibility to exonuclease. This susceptibility would be countered by Gam. It was postulated that Red-dependent recombination could generate rolling circles by creating a replication fork through strand transfer, and the consequent production of concatemeric DNA, the substrate required for the packaging of DNA to produce viable progeny. Thus, the idea emerged in 1974 that recombination could generate a replication origin.

Replication, recombination and Chi

Wild-type λ is a poor substrate for the host recombination system, but simple mutations were shown to convert it into an effective substrate. These mutations were isolated fortuitously in 1970 when it was discovered that a *red*$^{-}$ *gam*$^{-}$ λ can barely propagate. At this time, it was not known that both *red* and *gam* gene products contribute to the replication of the λ genome. Fortunately, the disadvantaged phage readily 'selects' mutations that help it to replicate. Each of these mutations was found to stimulate recombination of λ by the host system; they identified Chi (the cross-over hot-spot instigator), and made λ fundamental to the study of the bacterial recombination pathway. Chi defined an octanucleotide sequence, absent in the wild-type λ, but occurring around once per 5 kb in the *E. coli* chromosome. How, or when, Chi would have been discovered in the absence of λ is difficult to estimate.

Decades of elegant experiments documented the roles of double-strand breaks in DNA,

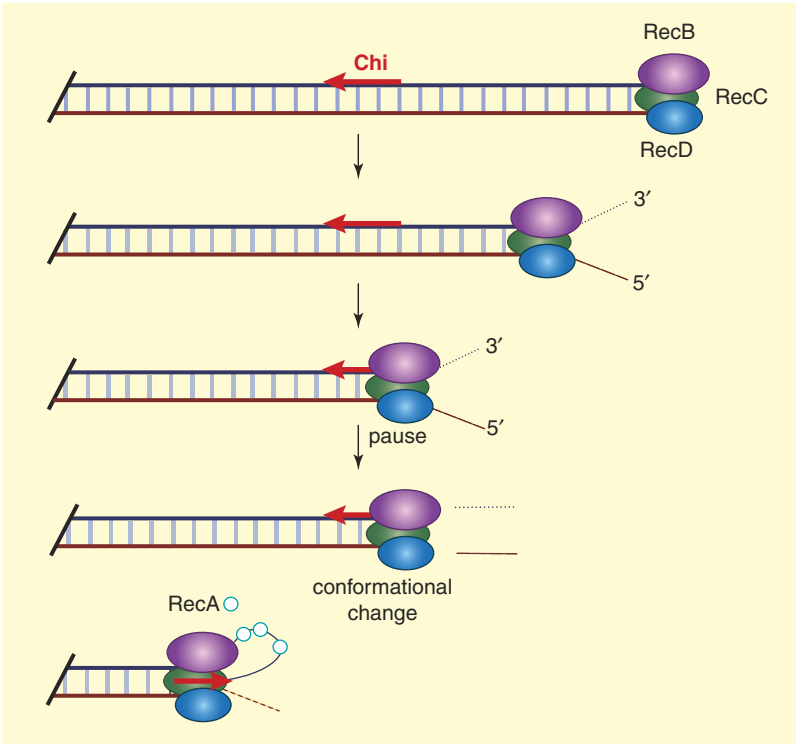

Figure 3. The role of Chi in recombination.

When RecBCD encounters a Chi sequence in the appropriate orientation, its conformation and, consequently, its activity change. Degradation now favours the 5′ end. RecA is loaded onto the conserved single strand, making this strand available for invasion of homologous double-stranded DNA and the initiation of recombination.

DNA replication and Chi sequences in recombination pathways.
In summary: double-strand breaks, the basic requirement for recombination, are commonly generated during DNA replication. Experiments with phages (λ and T4), *E. coli* and yeast subsequently documented this inseparable relationship between replication and recombination. For λ, recombinants can be generated in the absence of much DNA replication because a terminus of the λ genome (the right-hand) mimics a DNA break and serves as an entry point for either the Red or RecBCD enzymes. Chi sequences stimulate genetic exchange by the RecBCD pathway.

Biochemical experiments have demonstrated that Chi is critical for the role of RecBCD. This complex, a helicase with two motors, enters DNA at a double-stranded end (Figure 3) and degrades both strands of DNA until it encounters a Chi sequence in the 3′-terminated strand. When Chi is recognised, the activities of the protein complex change – the rate of translocation decreases; degradation of the 5′ strand is now favoured; and the modified enzyme becomes capable of loading RecA onto the single-stranded 3′ overhang, making that single-stranded DNA competent to initiate strand exchange. Recent experiments on single, trapped, λ DNA molecules are consistent with a conformational change in response to Chi which prevents the threading of the 3′-terminated strand into the relevant nucleolytic site and decreases the translocation speed. It has been proposed that the rate of DNA translocation is regulated to "co-ordinate the loading of RecA protein onto single-stranded DNA and the subsequent DNA pairing step of homologous recombination."

While there is no evidence for analogous sequences in eukaryotes, Chi has been fundamental to our understanding of recombination in bacteria.

Acquisition and loss of genes from genomes

The Campbell model, published in 1962, proposed that in a lysogen, the λ genome (the prophage) is an integral part of the bacterial chromosome. Today, it is hard for us to imagine any alternative, and consequently it is difficult to appreciate the impact of this model. Previously, disruption of the bacterial genome by insertion of the phage chromosome had been overlooked in favour of a model proposing side-by-side synapsis of the phage and bacterial genomes. Integration and excision of the λ genome led to the identification of site-specific recombination, while aberrant excision demonstrated an 'illegitimate' recombination pathway.

When the Campbell model was proposed, the λ chromosome was known to be linear. However, the circular nature of the linkage maps for *E. coli* and phage T4 were emerging. Furthermore, the gene order in the λ chromosome had been shown to differ from that in the prophage, and, if the phage chromosome were circular, a single reciprocal cross-over between specific sites within the phage and bacterial chromosomes would alter the gene order. Soon after this proposal, the ends of the λ genome were shown to be capable of cohering, providing a mechanism whereby the phage chromosome could circularize, prior to integration. The original genomes could be regenerated by the reverse genetic exchange, called excision.

Site-specific recombination

Integration of the λ genome into the host chromosome introduced the concept of site-specific recombination, so named because integration occurs at a specific site within both the phage and host genomes. The reaction is catalysed by a phage-encoded enzyme, but efficiency and accuracy depend on host factors.

Site-specific recombination is widespread in biology. It accounts for the movement of transposons and other mobile genetic elements around the genome — producing a lot of the spontaneous mutations found in many organisms. Half the human genome is made up of sequences derived from mobile genetic elements. Site-specific recombination is also vital for the segregation of circular

bacterial chromosomes, and in some specialized cases of gene expression. Moreover, phage P1, like λ, has a site-specific recombination system, called Cre-Lox, now widely used in experimental gene manipulation.

Lambda integration was the system in which much of the basic mechanistic underpinnings of site-specific recombination were worked out, from the early genetic studies through to recent structural examination of many of the steps in the process.

Integration works in outline as follows. The specific sites within the phage and bacterial genomes where recombination takes place during integration are called *attP* and *attB*, respectively. These sites each contain a similar core region containing binding sites for the phage-encoded enzyme, Integrase, which catalyses the strand-breaking and rejoining needed for integration. In addition, two other features of the system are critical. First, there are other binding sites for integrase outside the core *attP* site; second, another protein, the host protein Integration Host Factor (IHF), also binds nearby. Together these features ensure efficiency and also drive direction of the reaction.

Integrase has two DNA-binding domains, enabling it to bind simultaneously to one site outside the core sequence — in the so-called flanking sequences — and another within the core (where it acts). The flanking sites have higher affinity than those in the core region, and the high affinity sites help binding to the critical core sites. But filling of the core sites also requires binding of IHF, which, when it binds its sites, bends the DNA locally. Thus, IHF binds sites in the flanking regions of *attP*, and the bending of the DNA brings the strong integrase site in the flanking regions close to the weak site in the core, increasing filling of that core site.

Once the integration reaction has taken place, the *attP* and *attB* sites no longer exist: rather, there are now two new hybrid sites found at the boundaries between the phage and bacterial chromosomes. These are called *attR* and *attL* (for right and left).

Excision of the phage genome from the bacterial chromosome relies on the presence of another phage product — the DNA-binding protein Xis. This protein is made only under conditions of phage induction, and binds to sites in *attR* (originally in *attP*). As with IHF, Xis bends DNA when it binds. Together with IHF and integrase, Xis forms a protein–DNA complex at *attR* which interacts with another at *attL*, promoting recombination between these two sites, and restoring *attP* and *attB* (and the two separate DNA molecules, the phage and bacterial chromosomes).

Thus, we see how so-called architectural proteins, IHF and Xis, impart efficiency and directionality to the reaction. The role of IHF here was the first well characterized example of a DNA-binding protein functioning through DNA bending. Since then, IHF has been shown to work in other contexts, for example, helping bring together distantly bound regulators and the genes they control. Other architectural proteins have been identified that help gene activation in eukaryotes — for example, the HMG proteins that bind and bend the enhancer of the human interferon-β gene.

Illegitimate recombination and gene cloning by nature

While site-specific recombination regenerates the bacterial and phage genomes, a rare aberrant excision, resulting from a cross-over between a site within the bacterial genome and one within the phage genome, can produce a phage that has acquired bacterial genes. Such phages, so-called transducing phages, primarily λ*gal*, were recognised in the 1950s by their ability to transfer genes between bacteria.

The genetic analysis of λ*gal* phages confirmed that each acquired a segment of bacterial DNA at the expense of phage genes. The lengths of the DNA segments, lost or gained, were variable. These findings, in addition to their support for the Campbell model, identified recombination between nonhomologous DNA. This process of 'illegitimate recombination' requires only short segments of sequence identity (5–14 base pairs) and occurs with low frequency, but is now believed to be of general significance to the evolution of new genomes through the acquisition, loss and exchange of DNA segments (Box 2). The sequencing of many bacterial genomes has highlighted the role of lateral DNA transfer in the evolution of those organisms.

Lambda and genetic engineering

The Campbell model explains how a λ genome can acquire genes from the bacterial chromosome, and how homologous recombination can exchange mutations between incoming bacterial genes carried by a transducing phage and the bacterial chromosome. λ*gal* phages, mentioned above, were the pioneers in this story. Tricks were subsequently used to get λ to integrate close to genes of interest so that the phage could acquire them. This, for example, allowed cloning of the *lac* and *trp* operons *in vivo*.

Before the advent of recombinant DNA technology, transducing phages demonstrated the potential of gene cloning. In the 1960s the analysis of gene expression using λ and the *lac* operon as substrates were pursued at a molecular level. It became obvious to those working with the *lac* system that it was easier to study the interaction of the λ repressor with its operator sequences within the λ genome of 49 kb than the Lac repressor with its operator sequences within the bacterial genome (4,700 kb). λ*lac* transducing phages were derived *in vivo* and they illustrated many benefits of cloned DNA — for example, the enrichment and amplification of a DNA sequence to serve as a substrate for binding proteins and, in 1968, 100-fold over-production of Lac repressor.

Early transducing phage were used to explore expression of cloned genes. Two features of λ exemplified general principles for the amplification of a protein. First, the genome can be maintained in single copy and induced to replicate to provide many copies when gene expression is required. And second, the cloned gene can be transcribed from a strong, controllable promoter. Recent 'copy control' plasmids (pETcoco) emulate these features.

Box 2

Illegitimate recombination, horizontal gene transfer and evolution.

Comparisons of the λ genome with those of other λ-like phages were reported in 1971 following the examination of DNA heteroduplexes by electron microscopy. These analyses revealed that the genomes of related phages were 'mosaic': that is, the genomes were modular, consistent with alternative modules being reshuffled during evolution. It was proposed that genetic exchange of modules occurred within short linker regions. The current availability of genome sequences has enhanced our appreciation of mosaic viral chromosomes. Extensive, and broadly based, comparative analyses of genome sequences support the importance of horizontal exchange of segments of DNA as a key feature of viral and microbial evolution. Currently, the role of illegitimate recombination is emphasized. As shown by the creation of λ*gal* phages, illegitimate recombination generates novel combinations of genes. Although the majority of hybrid phage generated in this way are unlikely to survive, some will, and their genomes will be refined by homologous recombination and deletions.

Current evidence of divergence times derived from DNA sequence comparisons also suggests that bacteriophages may have ancestral connections with the viruses of eukaryotes and archaea. Illegitimate recombination, transduction and transposition are all likely to contribute to horizontal gene transfer.

Genetic engineering by natural processes quickly enhanced the molecular, biochemical and genetic analysis of *E. coli*. By 1968, the second edition of Bill Hayes' classic textbook already diagrammed the exchange of mutations between a bacterial gene in a λ transducing phage and the bacterial chromosome. Two cross-overs are required, one on either side of the genetic difference. Using a transducing phage encoding a temperature-sensitive repressor, selection of the first of the two cross-overs is achieved by isolating lysogenic cells — these are selected at low temperature when the functional repressor makes the lysogen immune to super-infection. Selection for a second cross-over depends on the subsequent survival of cured cells at high temperatures, which only cured cells can do. Some of the cured cells will have been generated by a cross-over on the appropriate side of the genetic difference.

The use of λ vectors to clone DNA fragments generated *in vitro* was reported in 1974. Cloned genes could be manipulated and returned to the bacterial chromosome. Almost as soon as yeast was made competent in the uptake of DNA, the Campbell model influenced the design of a system to transfer genes from a yeast replicon to the yeast chromosome. The transfer of manipulated genes by homologous or site-specific recombination has become fundamental to the analysis of many microbial and eukaryotic organisms.

Studies of λ morphogenesis led to the means of *in vitro* packaging of λ genomes. Initially this was applied to recombinant λ molecules, but subsequently to cosmids and phage P1. *In vitro* packaging had a major impact on recombinant DNA technology by enhancing the efficiency of construction of both genomic and cDNA libraries.

Recombineering

From our knowledge of phage recombination systems, new technologies (recombineering) have emerged for making novel combinations of DNA in *E. coli* by recombination, rather than *in vitro* by restriction enzymes and DNA ligase. These reactions rely on the λ gene product Gam to inhibit the exonuclease activity of RecBCD. In the presence of Gam the linear DNA generated by PCR can serve to donate a DNA segment to a replicon in *E. coli* — transfer by the phage recombination system (Redα and Redβ) requires only short regions of homology (<50 base pairs). This permits the construction of sophisticated gene fusions (e.g. the insertion of markers, regulatory elements, or sequences to aid in the purification of the protein product), while mutations can be made in a cloned gene using Redβ to promote assimilation of an oligonucleotide within a replication fork. The *in vivo* manipulation of DNA in *E. coli* emulates the methods used in yeast, but it offers the important advantage of enhanced yields of DNA. Changes made to mouse DNA cloned in a BAC vector in *E. coli* become available for transfer to the mouse genome. The use of phage and viral recombination systems has recently been extended to other bacteria and some eukaryotes.

Acknowledgments

We wish to acknowledge the discoverer of lambda, Esther Ledeberg, who died in November 2006. We also thank Allan Campbell, John Connelly, Don Court, Richard Ebright, John Ellis, Sydney Gary, Ken Murray, Mark Ptashne, Frank Stahl and Bob Weisberg for help with the manuscript.

Further reading

Atsumi, S., and Little, J.W. (2006). A synthetic phage λ regulatory circuit. Proc. Natl. Acad. Sci. USA *103*, 19045–19050.

Campbell, A. (2003). The future of bacteriophage biology. Nat. Rev. Genet. *4*, 471–477.

Court, D.L., Sawitzke, J.A., and Thomason, L.C. (2002). Genetic engineering using homologous recombination. Annu. Rev. Genet. *36*, 361–388.

Dodd, I.B., Shearwin, K.E., and Egan J.B. (2005). Revisited gene regulation in bacteriophage λ. Curr. Opin. Genet. Dev. *15*, 145–152.

Ellis, R.J. (2005). Chaperomics: in vivo GroEL function defined. Curr. Biol. *15*, R661–R663.

Gottesman, M.E., and Weisberg, R.A. (2004). Little λ, who made thee? Microbiol. Mol. Biol. Rev. *68*, 796–813.

Grindley, N.D., Whiteson, K.L and Rice, P.A. (2006). Mechanisms of site-specific recombination. Annu. Rev. Biochem. *75*, 567–605.

Handa, N., Bianco, P.R., Baskin., R.J., and Kowalczykowski., S.C. (2005). Direct visualization of RecBCD movement reveals cotranslocation of the RecD motor after Chi recognition. Mol. Cell *17*, 745–750.

Hendrix, R.W. (2002). Bacteriophages: evolution of the majority. Theor. Pop. Biol. *61*, 471–480.

Murray, N.E. (2006). The impact of phage λ: from restriction to recombineering. Biochem. Soc. Trans. *34*, 203–207.

Ptashne, M. (2004). A Genetic Switch: Phage Lamba Revisited, Third Edition (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Stahl, F.W. (2005). Chi: a little sequence controls a big enzyme. Genetics *170*, 487–493.

[1]Institute of Cell Biology, The University of Edinburgh, Edinburgh EH9 3JR, UK.
[2]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.
E-mail: [1]noreen.murray@ed.ac.uk, [2]ganna@cshl.edu