Molecular recognition

Feeling for the bumps

Robert Schlief

THE problem of molecular recognition faced by the proteins that decide the fate of a bacterium infected with phage lambda is not unlike the problem of identifying a human in a dark room and then distinguishing a man from a woman. The two possible fates of the bacterium are lysis, if the phage expresses its lysis genes and multiplies until it lyses the cell; or lysogeny, if it expresses an alternative set of genes that enable it to integrate into the bacterial chromosome and replicate along with it until the lytic programme is reactivated by environmental insult. The decision between lysis and lysogeny hinges on the differential recognition by two phage proteins of operator sites in the phage DNA. Binding to these sites determines which phage proteins will not be expressed, and hence the developmental pathway followed by the phage. By a combination of reverse genetics and biochemistry, A. Hochschild and M. Ptashne and collaborators (*Cell* **44**, 925–933 and **45**, 681–687; 1986) have been able to show exactly how the fine discrimination required for differential binding is achieved.

The two proteins that have this pivotal role in the decision between lysis and lysogeny are Cro and repressor. Each is capable of binding to three closely related operator sites, O_R1, O_R2 and O_R3. In fact repressor does bind to all three operators at times in a stably lysogenized cell, and Cro does so late in the lytic cycle. The figure shows how the binding of Cro and repressor to these sites in a newly infected cell determines the decision between lysis and lysogeny. The crucial point is that as the proteins begin to accumulate in the cell they do not bind to the three sites in the same order: Cro has the highest affinity for O_R3 and first binds there, whereas repressor first binds to O_R1. If Cro takes the ascendency in an infected cell, its binding first to O_R3 stops the synthesis of repressor but allows the transcription of the lysis genes. Alternatively, the binding of repressor first to O_R1 would halt the synthesis of Cro and lead to establishment of lysogeny. Thus, the choice of genetic programme followed by the phage rests on Cro and repressor proteins being able to both recognize and distinguish O_R1 and O_R3.

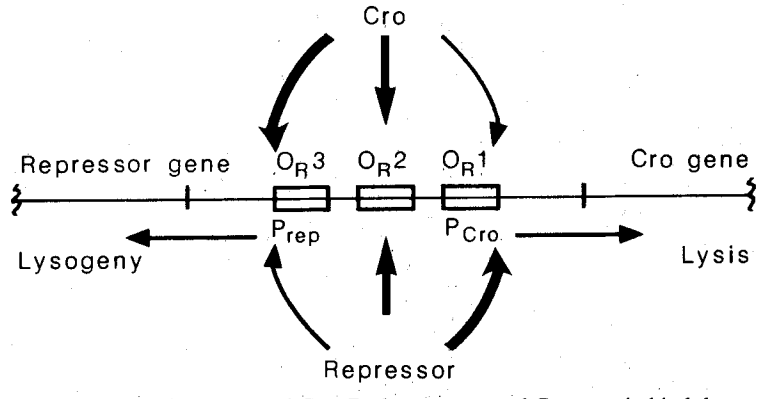
The binding properties of Cro and repressor raise two questions. First, are the Cro and repressor binding sites exactly centred upon each other, or do they merely share a few nucleotides in common? Then, if the binding sites directly overlap, exactly how do Cro and repressor tell O_R1 and O_R3 apart?

The first question is simply answered. Cro and repressor do indeed contact DNA homologously. Both proteins contact DNA primarily via a protruding alpha helix that neatly fits into the major groove of the DNA. Amino-acid residues in common between the helices of the two proteins contact common structural elements of the operators. For example, the first of the two sets of experiments from Ptashne's laboratory shows that one of these, a serine residue located at the second position in the helix of both proteins, contacts position four of O_R1 and O_R3. This result eliminates the possibility of separate but slightly overlapping binding sites for Cro and repressor. The binding sites of the proteins exactly overlap.

The second question — how the sites are distinguished by the proteins — must therefore be answered. The most reasonable possibility, and the one which is reported in the second paper, is that the proteins directly read the three-nucleotide differences between the sites. For example, Cro detects the presence of a thymine at position three of the operator but repressor is constructed to be indifferent to the base at position three. Analogously, Cro is indifferent to the base at position eight whereas repressor contacts this position and binds significantly more tightly if it is a guanine, as is found in O_R1. As a final demonstration, Hochschild, Ptashne and collaborators identified the amino-acid residues contacting the three nucleotides that distinguish O_R1 from O_R3 and showed that placing the Cro amino acids in repressor shifted the repressor specificity towards that of Cro and vice versa.

To demonstrate the indifference of Cro to the identity of base eight in the operators, Hochschild *et al.* synthesized operators of suitably altered sequence and measured their affinity for Cro and repressor by DNase footprinting. The demonstration that a specific residue of the proteins contacts a specific base used much the same idea. First, the affinity of the wild-type protein was shown to be sensitive to the identity of a specific base of the operator. Second, when a smaller amino-acid residue was substituted for the original residue at the position of hypothesized contact, the affinity of the altered protein became indifferent to the identity of the base in that particular position. This is as expected, for the smaller amino acid fails to contact the DNA, and therefore the binding energy of the protein is independent of the base at the position normally contacted by the substituted amino-acid residue. This type of approach has also apparently been successful in demonstrating a specific amino-acid-residue–base contact between *lac* repressor and *lac* operator (Ebright, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 303–307; 1986).

What about the future? In few other systems will we have available even the relatively coarse X-ray structure data of lambda repressor bound to DNA (Anderson, J. *et al. Nature* **316**, 596–601; 1985), vast amounts of biochemical data, results from nuclear magnetic resonance experiments and a rich biology, all of which



Operator sites binding repressor and Cro. Both repressor and Cro protein bind three related operators, O_R1, O_R2 and O_R3, but with inverse affinities (schematically indicated in the diagram by the relative thicknesses of the arrows): repressor binds O_R1 most strongly and Cro binds O_R3 most strongly. In a lysogenic bacterium, repressor is bound at O_R1 and at O_R2: binding of repressor at these two sites, which is cooperative, prevents transcription of Cro and other genes to the right that are necessary for the lytic cycle, but activates transcription of the repressor gene from its promotor, P_{rep}, to the left: thus repressor is continuously synthesized in concentrations high enough to enable it to bind to O_R3, for which its affinity is relatively weak; and lysogeny is maintained. The lytic cycle is activated when an environmental insult such as ultraviolet radiation induces a protease that cleaves the repressor protein in such a way that cooperative binding to O_R1 and O_R2 is no longer possible, substantially reducing the affinity of repressor for O_R1 and O_R2. With the release of repressor from O_R1 and O_R2, Cro synthesis can begin from its promoter P_{Cro}. The Cro protein then binds to O_R3 and shuts off transcription of repressor so that expression of the other lysis genes is no longer prevented, and the lytic programme is initiated.