SMR/111 - 20

SECOND SUMMER COLLEGE IN BIOPHYSICS

30 July - 7 September 1984

STRUCTURAL STUDIES OF GENE REGULATION IN THE BACTERIOPHAGE λ SYSTEM

I. Introduction to Biological Information

II. Introduction to λ Genetics

III. Introduction to λ Repressor and λ Cro Protein: the Helix-Turn-Helix Class of DNA-binding Proteins

IV-V. NMR Studies of λ Repressor and λ Operator

M. WEISS
Department of Chemistry
Harvard University
Cambridge, Massachusetts 02138
U.S.A.

Michael Weiss
August, 1984

Lecture 1

Introduction. The purpose of this set of 5 lectures is to describe attempts to understand the molecular basis of protein-DNA interactions. The sequence-specific recognition of DNA by proteins underlies the fundamental biological process of information control, and so in these lectures we will begin with biology, then to aspects of genetics and biochemistry, and finally to biophysics. The bacteriophage λ will be used as a model system at each level. Our unifying theme through the course of these lectures will be the nature and control of information.

I. Information: Information — its origins, evolution, regulation and expression — lies at the heart of biology. We may classify biological information into 3 forms

A) extrinsic

B) intrinsic

C) meta (or system-level)

Meta Information has evolved as a system-property that is more than the sum of its parts. For example, when we write

DNA → RNA → Protein

We mean, ~~e.g.~~,

"dGAA" → "GAA" → "Glutamic Acid"

I use quotation marks to emphasize that these are linguistic relationships. There is nothing intrinsic in the chemistry of the stretch of bases dGAA that imbues it with the meaning glutamic acid. Rather, there is a higher-order grammar in the non-equilibrium chemistry of the cell that provides the meaning. Context implies content: the sentence "dGAA specifies glutamic acid" is a property of the system. The evolution of grammar is a deep and unsolved problem.

Extrinsic information is the meaning of an individual sentence according to the meta rules of the cell's grammar. For example,

5'... dCAG CAG TCT ATA GAG CAG CTC ... 3'

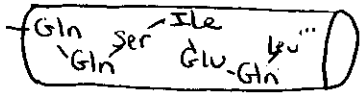
↓

N... Gln-Gln-Ser-Ile-Glu-Gln-Leu- ...

is a translation of the extrinsic information encoded in the string of symbols "CAGCAGTCT..." That the β-globin gene encodes the β-chain of hemoglobin is an example of the extrinsic information in our chromosomes.

Intrinsic information is contained in the biophysical chemistry of the biomolecules themselves. For example, the suggested folding of phage 434 cI repressor,

Gln – Gln – Ser – Ile – Glu – Gln – Leu – ...



α-helix

is an example of intrinsic information contained in the 1° sequence of a protein. There is no gene (string of base symbols) which encodes the sentence (whose meaning would be imbued by meta rules) "fold the portion of the repressor protein '... Gln-Gln-Ser-Ile-Glu-Gln-Leu-' into an α-helix." Indeed, α-helices are never mentioned in our genetic information explicitly. They are an intrinsic property of the way polypeptides fold in solution. The chemical principles which relate extrinsic and intrinsic information (protein folding, protein-DNA recognition, etc.) are the subject matter for biophysics.

II. Bacteriophage λ

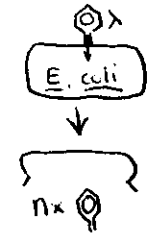
λ is a virus that infects the microorganism Escherishia coli. It is a simple (biologically speaking)

example of a life form with 2 states.

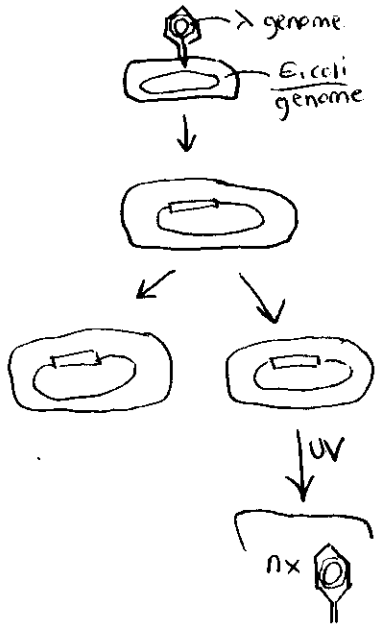
State I Lytic Growth

State II Lysogenic Growth

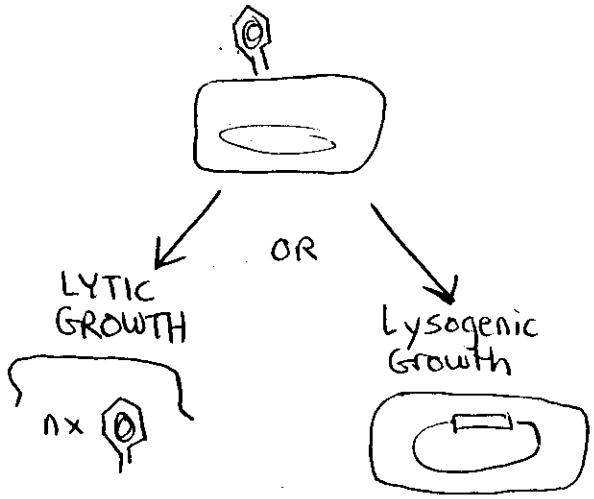
Lytic Growth means that the virus, after injecting its DNA into the bacterial cell, multiplies, kills its host and releases ~10^2 copies of itself into the environment.



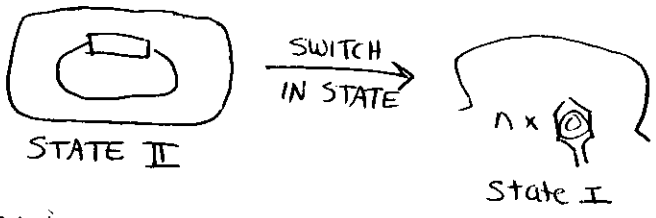
Lysogenic Growth means that the virus, after injecting its DNA into the cell, integrates its DNA genome into the bacterial genome, and becomes a quiescent passenger, a dormant parasite sequence of bases passed on to successive daughter cells in the normal course of bacterial replication. The dormant prophage can be induced into lytic growth by exposure of the bacterium to agents that damage DNA (e.g., UV light, certain mutagens).



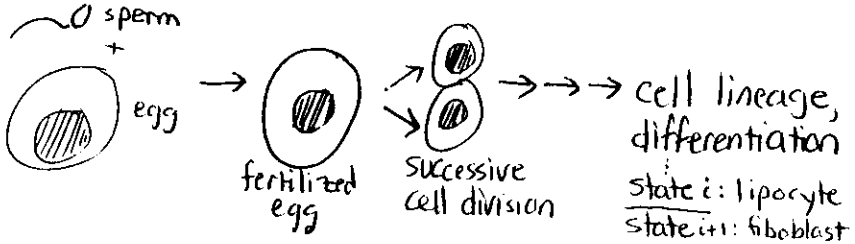
Thus, when the λ genome enters a cell, there is a developmental decision:



Likewise, exposure of the lysogenic cell to UV light occassions another (and related) decision:



This type of behavior represents a general biological phenomonon. For example,



Let us step back and view the λ life cycle in terms of information:

meta information

is borrowed from host bacterium. The virus consists of a stretch of duplex DNA surrounded by capsid proteins. It is metabolically inert. Without the meta information provided by the host, its DNA has no extrinsic information and is really not "alive."

extrinsic information

the sequence of bases in its genome as read by the host bacterium.

intrinsic information

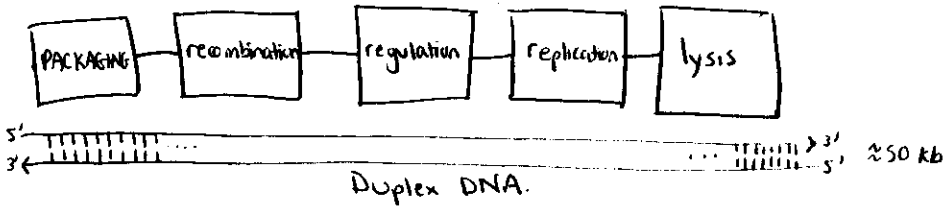
A) the biophysical properties of the phage head: recognition of bacterial wall, injection of naked DNA into cell cytoplasm

* B) the intrinsic properties of encoded proteins and DNA binding sites that allow replication, regulation, and assembly of progeny.

Sidney Brenner introduced the concept of genetic program in describing the development of an organism. Let us apply it to λ phage.

Bacteriophage λ

Genetic Map

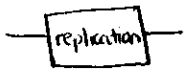


Duplex DNA.

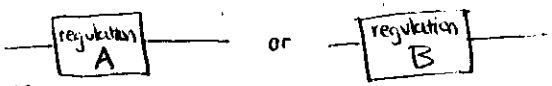
linear sequence of bases → functional subdivisions to form modular programming

Genetic Module = subroutine in genetic program

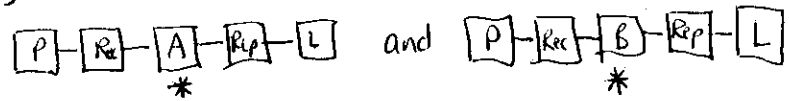
e.g.,



David Botstein has suggested that such modules are independent agents in evolution of lambdoid family of phages: λ, 434, P21, P22, ... That is, a phage may be viewed as an integrated collection of modules (subroutines). If several chemically distinct but functionally similar modules existed, then a new master program could be built by reshuffling. For example,

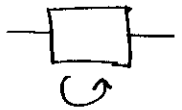


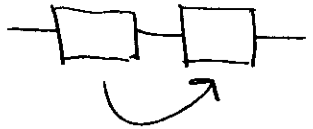
giving rise to different phages



These two phages may appear different, have diffent names, different host specificities, etc. Nevertheless, there is an underlying unity in the logic and higher-order architecture. This is a sort of "meta-species."

Clearly there are (at least) 2 levels of control,

A)  the logic within a module (intrinsic and extrinsic)

B)  the relationship between modules

This may be viewed as the master program which calls the various subroutines in a well-structured fashion.

We will now look in more detail at the "flow-chart" of λ development. The central question we will ask is — what is the nature of the hardware that makes possible this modular software?

Terminology

1. Operator

   sequence of DNA bases that form a binding site for proteins that regulate, positively or negatively, the initiation of transcription.

2. termination site

site of termination of transcription along sequence of DNA bases. Termination and anti-termination under control by proteins.

3. Regulatory Proteins

Proteins whose role is to control and coordinate levels of gene expression. This may occur at the level of transcription, mRNA processing, translation or post-translational modification. May also occur via control of DNA topology.

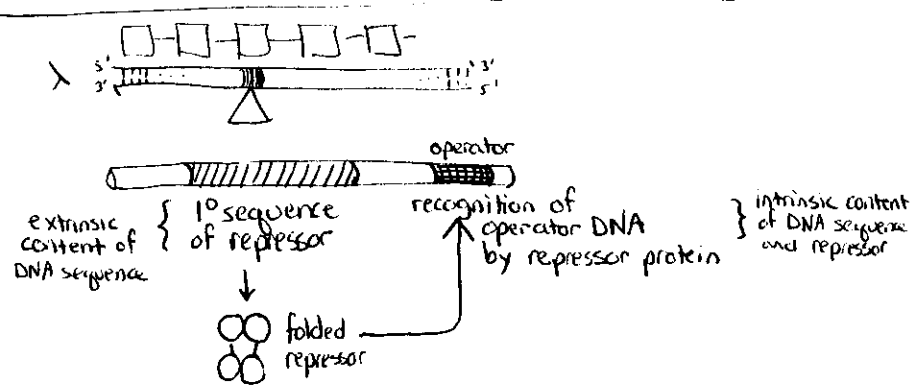
4. Promoter

sequence of DNA bases recognized by RNA-polymerase as start point for transcription.

5. Enhancers

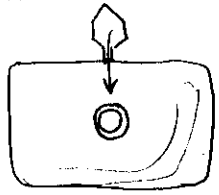
sequence of DNA bases in eukaryotic systems that influence levels of transcription, perhaps via sequence-dependent geometric/topologic properties, e.g., B → Z DNA. Not relevent to λ phage.

Intrinsic and Extrinsic Information in Master Program



Lytic Development

1. phage DNA enters cell



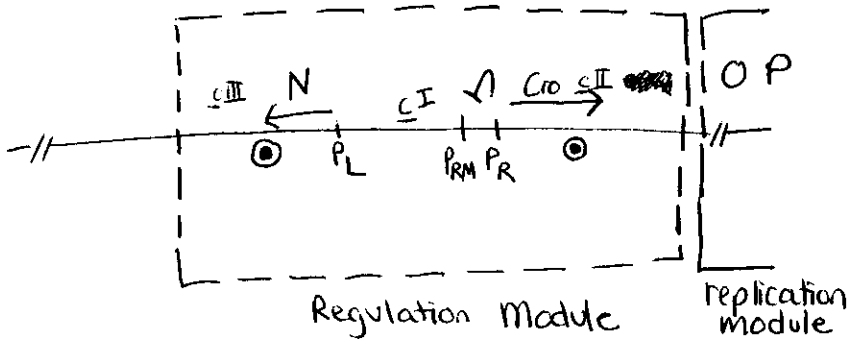
Notation

gene G

protein pG where p denotes gene product

2. Early Genes

A) Immediate Early Transcription



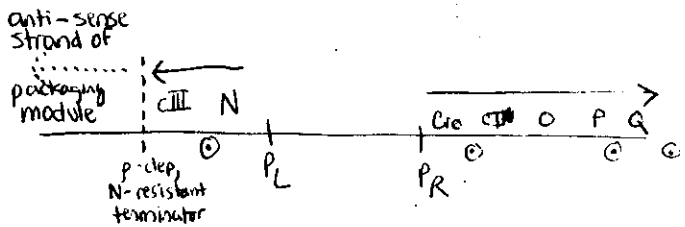
pCro — a repressor that binds to the operators O_L and O_R to negatively regulate initiation of transcription. Turns off P_{RM}-mediated transcription of cI.

pN — a regulatory protein which anti terminates transcription at [redacted] ⊙

B) Delayed Early Transcription

The N gene product anti-terminates, allowing read through of the P_L and P_R transcripts.

Delayed Early

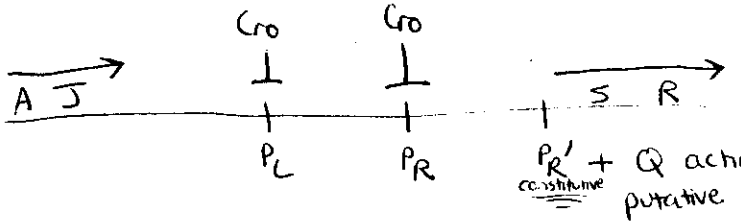


pO, pP — proteins involved in autonomous DNA replication

pQ — another anti terminator involved in subsequent control.

[pCro protein] builds up to repress P_L, P_R-mediated transcription.

3. Late gene expression



pS, R — lysis function

pA→J — packaging of progeny phage heads.

Lecture 2

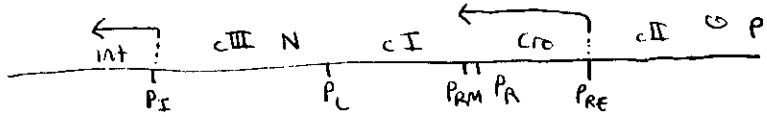
Lysogenic Growth and the Lytic-Lysogenic Switch

1. phage DNA enters cell

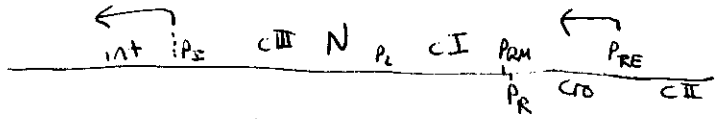
2. early gene expression as before
   A) immediate early → pN, pCro
   B) delayed early
      → pcII *
      pcIII *

Action of pcII

1. activates transcription promoters P_{RE} P_I



Remember that cIe is read rightward from P_R, and so the P_{RE} transcript reads the anti-sense strand of cro before reading the sense strand of cI. To make this clear, we will rewrite the genome with rightward-read genes below, leftward-read genes above:



2. retards lytic growth
   a) inhibits DNA synthesis
   b) reduces Q expression. Hypothetical mechanism
      right-ward transcription from P_R (to make Q) slowed by concomitant leftward transcription from P_{RE}.

p cII is highly unstable and is stabilized by p cIII.

Next step in lysogenic pathway:

3. p cII, c III – mediated expression of int and c I.

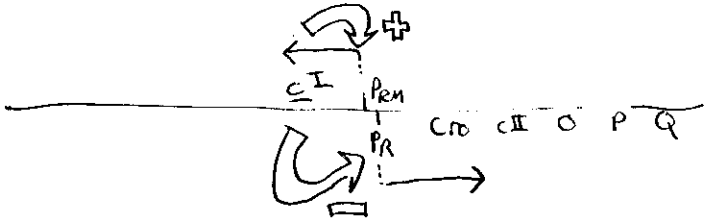
p int → site-specific recombination into bacterial genome.

c I = λ repressor

a) turns off P_R and P_L

Note that Cro is off and Q is off.

b) Stimulates transcription from P_{RM} to make more c I (positive feedback)



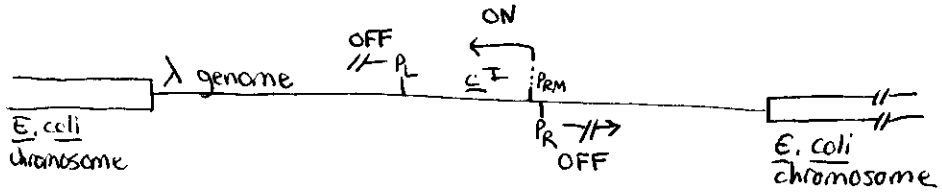
Positive Feedback

p c I stimulates its own transcription in lysogeny

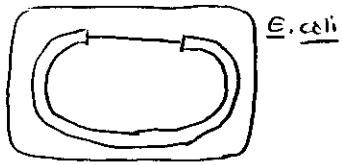
Negative Feedback

p Cro inhibits its own transcription in lytic program.

Lysogenic State



Switch: Lysogenic → Lytic



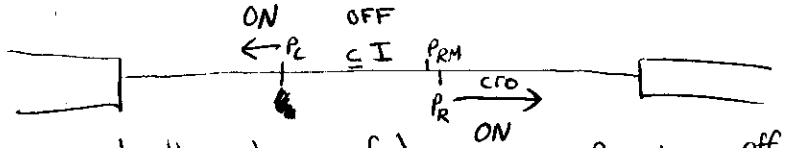
1 ↓ exposure to UV or class of mutagens

2 ↓ DNA damage

3 ↓ induction of "SOS" host response

4 ↓ inactivation of λ repressor (see below)

* SWITCH IN STATE



In the absence of λ repressor, P_{RM} turns off and P_R is free from repression and turns on. p Cro then keeps c I off and only at a later time turns itself off as well, as in the lytic pathway.

Through this point we have discussed phenomenology without understanding its structural basis. We should appreciate fully how the science of genetics (in and of itself) can give a clear phenomenological picture of the key players and their relationships

1) recombination mapping gives a linear ordering of genetic activities without DNA sequencing.

2) analysis of mutations, where they lie, their complementation groups etc yield a functional map.

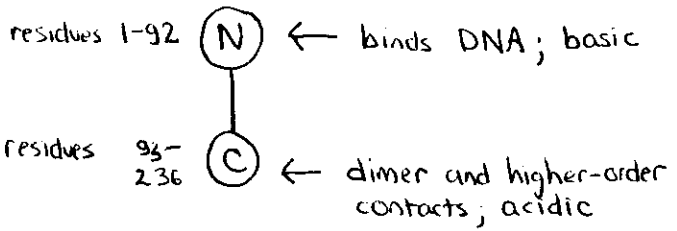
3) comparative genetics of lambdoid family of phages and hybrid phages give insight into modular programming.

We most appreciate that 50 years of genetic analysis underlies all subsequent biochemical and biophysical studies.

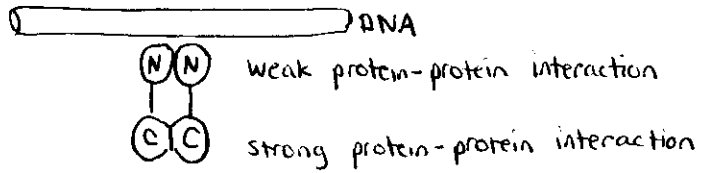
We may now turn our attention to the molecular hardware of the λ system.

λ Repressor

2-domain protein

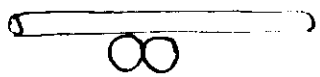


binds to DNA as a dimer



λ Cro Protein

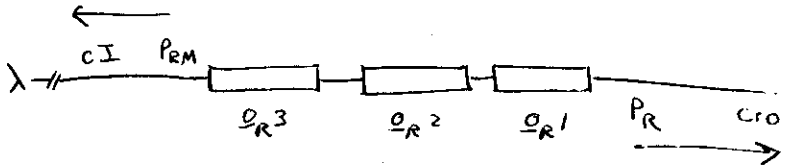
66 amino acids
1 domain
basic
strong dimer



λ Operator Sites

DNA binding sites for repressor and cro protein.
17 base-pairs long
almost but not quite symmetric about central base-pair
Operator = set of 3 operator sites

For example, the right operator O_R



Right Operator contains 3 operator sites O_R1, O_R2, O_R3. These are of related (but not identical) DNA sequence, which is almost (but not quite) symmetric about the central base pair.

O_R3 5' TATCACCGCAAGGGATA 3'
3' ATAGTGGCGTTCCCTAT 5'

O_R2 5' TAACACCGTGCGTGTTG 3'
3' ATTGTGGCACGCACAAC 5'

O_R1 5' TACCTCTGGCGGTGATA 3'
3' ATGGAGACCGCCACTAT 5'

[] = point of sequence symmetry

There are 6 bp between O_R3 and O_R2

5' AATATC 3'
3' TTATAG 5'

and 7 bp between O_R2 and O_R1

5' ACTATTT 3'
3' TGATAAA 5'

Various base-pair substitutions in the operator site can increase, decrease, or have little effect upon repressor and Cro binding affinity; models of their respective protein–DNA complex must account for these data.

① perfectly symmetric sequence

5' TATCACCGCGGTGATA 3'
3' ATAGTGGCGCCACTAT 5'

binds better than wild-type sites.

② departures from consensus sequence required for biological function, as we shall see. Optimal binding affinity ≠ optimal biological function

③ the repressor can also be "improved" to bind better, but this would also be a biologically "poor" molecule.

Pattern of Repressor and Cro Binding to O_R

A. Cro

1. Cro binds most tightly to $\underline{O}_R3$ and only at higher concentrations to $\underline{O}_R2$ and $\underline{O}_R1$.

2. Binding to 3 sites is non-cooperative.

3. Binding to $\underline{O}_R3$ turns off P_{RM} without affecting P_R. Thus, at low cro concentrations (early lytic pathway)

   P_{RM} off → no cI repressor
   P_R ON → transcription of Cro, cII, ...

4. Binding to $\underline{O}_R2$ and $\underline{O}_R1$ at higher Cro concentrations (late lytic pathway)

   P_{RM} off (as above)
   P_R off → Cro turns itself off, (and rest of P_R operon off).

B. λ cI Repressor

1. cI repressor binds most tightly to $\underline{O}_R1$, then $\underline{O}_R2$ (~2x↓), then $\underline{O}_R3$ (~10x↓).

2. Binding is pairwise cooperative. As a result, effective binding affinity

   $$\underline{O}_R1 \approx \underline{O}_R2 > \underline{O}_R3$$

   Thus, order is reverse of Cro.

3. Binding of repressor to $\underline{O}_R1$ and $\underline{O}_R2$
   a) turns on P_{RM}
   b) turns off P_R

Binding to $\underline{O}_R3$ at higher concentrations, which does not happen physiologically, turns off P_{RM} also. In lysogen, $\underline{O}_R3$ only ≈ 20% occupied.

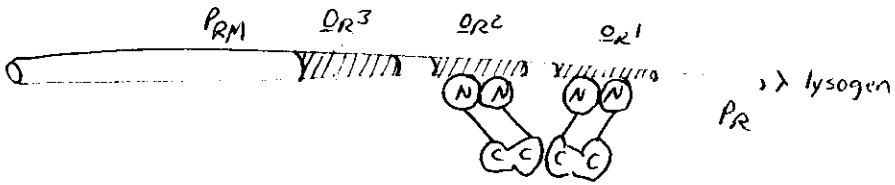
Genetics

	P_{RM}	P_R
Lysogenic State	ON	OFF
Lytic State	OFF	early ON, late OFF

Biochemistry

	O_R3	O_R2	O_R1
Lysogenic State	80% VACANT	cI	cI
Lytic early	Cro	—	—
Lytic late	Cro	Cro	Cro

The Lysogenic → Lytic Switch (see Ptashne review)



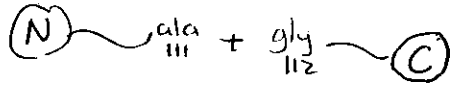
↓ UV
induction of SOS response

host rec A protein binds to C-domain

↓ conformational change*

autocleavage of cI repressor
— ala 111 ↓ gly 112 —

to yield



*Little et al PNAS USA (1984)
Gimble and Sauer (unpublished)

The binding affinity of the isolated N-terminal domain is much lower than the intact protein, and so

① P_{RM} positive control diminishes in absence of intact cI → less repressor still.

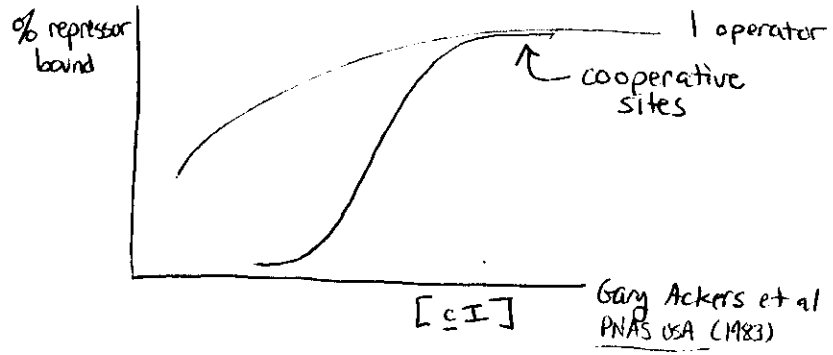
transient signal * ② basal transcription of P_R now unrepressed by repressor → synthesis of Cro

③ Cro binds to O_{R3}, maintains repressor off, and lytic program begins.

Thus, a transient signal allows a switch between two stable states. Nevertheless, the system is not "leaky": in absence of SOS-mediated induction, spontaneous switch occurs only 1/10⁵.

How can you have a System that is both stable and easily switched?

This looks at first glance like contradictory goals. It is accomplished by cooperativity of repressor binding to adjacent operator sites.



Gary Ackers et al PNAS USA (1983)

Biological Constraints of Repressor-Operator Interaction

① If repressor bound <u>too</u> well, system would not be easily inducible. A "super repressor" would not conform to the requirements of the phage <u>software</u>.

② promoter and operator sequences overlap. Optimal promoter sequence ≠ optimal operator sequence as a general rule.

③ Optimal sequence ~~for~~ for Cro ≠ optimal sequence for repressor. The 2 respective affinities must be balanced and independently perturbed to establish opposite orders of affinity for sites 1, 2, 3.

④ Cro merely has to bind to site to turn <u>off</u> adjacent promoter. Repressor has the extra job of turning <u>on</u> P_{RM}. This constrains the amino-acid sequence of repressor. For example, the mutation

Glu 34 → Lys

in helix 2 of repressor enhances DNA affinity*, but eliminates positive control of P_{RM}.

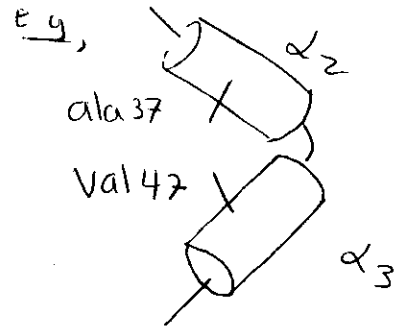
*Hochschild et al
<u>Cell</u> 32: 319 (1983)
Nelson, H.C.M. (personal communication)

Lecture 3

Topic 1: Helix-Turn-Helix Motif

A. Crystal Structures
1. <u>c</u>I
2. Cro
3. CAP

B. Supersecondary and tertiary interactions



C. Scaffold for Recognition
1. λ repressor genetics
2. helix-switching recombinants
3. altered-specificity mutants

D. The N-terminal Arm of λ Repressor

Lecture 4 and 5

NMR Studies of λ System

1. Introduction to two-dimensional techniques.
2. Use of molecular biology techniques in NMR studies: mutants as structural tools
3. NMR of DNA
4. NMR characterization of λ repressor
5. Structure - Function Relationships
   - A. N-terminal Arm
   - B. Dimerization α_5
   - C. Helix-Turn-Helix Motif
6. Conclusions and Future Prospects

References (lectures 1 and 2)

1. Lambda volume (1970) Cold Spring Harbor Press, Ny
2. Lambda II volume (1983) Cold Spring Harbor Press, NY
   esp.
   Chapter 2, "Lytic Mode" D. I. Friedman & M. Gottesman
   Chapter 3, "Establishment of Repressor Synthesis" D. L. Wulff
   Chapter 5, "Repressor and Cro Protein" G. N. Godson & M. Rosenberg, A. D. Johnson, C. O. Pabo, R. T. Sauer
3. Botstein, D. and I. Herskowitz, Nature 251: 584 (1974) Properties of hybrids between Salmonella phage P22 and coliphage λ
4. A. D. Johnson et al. Nature 294: 217 (1981) λ Repressor and cro — components of an efficient molecular switch
5. Mark Ptashne et al. "A Genetic Switch in a Bacterial Virus" Scientific American (11/82) 247: 128
6. Maurer, R. et al. J Mol Biol 139: 147 (1980)
7. Meyer, B. et al., J Mol Biol 139: 163 (1980)
8. Meyer, B. et al, J. Mol Biol 139: 195
9. Johnson, A., et al Proc Natl Acad Sci USA 75: 1783 (1978)

   } describe operation of λ operators

10. Hofstadter, D. Goedel, Escher, Bach: the eternal golden Braid chapter on DNA coding and information theory.

References II (lecture 3)

1. Primary Sequence of λ Repressor

Sauer R.T. and Anderegg, R *Biochemistry* 17 1092 (1978)

2. Crystal Structure of λ Repressor.

Pabo, C.O. and Lewis, M. *Nature* 298 443 (1982)

3. N-terminal Arm

Pabo, C.O., Krovatin, W., Jeffrey, A. and Sauer, R.T.
Nature 298: 441 (1982)

4. Crystal Structure of Cro

[redacted]
Anderson, W.F. et al *Nature* 290:754 (1981)

5. Crystal Structure of CAP

McKay, D.B. and Steitz, T.A. *Nature* 290: 744 (1981)

6. NMR Structure of *lac* Repressor

Zuiderweg et al *Proc Natl Acad Sci USA*
80: 5837 (1983)

7. Sequence Homologies

Sauer, R.T. et al *Nature* 298 :447 (1982)

8. Genetics

Hecht et al *Proc Natl Acad Sci USA* 80 :2676 (1983)
Hochschild et al *Cell* 32 : 319 (1983)
Warton et al *Cell* (in press)
Ebright et al *Nature* (in press)

9. Mechanism of Transcription - Kinetics

Hawley and McClure *Cell* 32 : 327 (1983)

References (lectures 4 and 5)

1. Genetics and NMR

Jarema, M.A.C. et al *Proc Natl Acad Sci USA* 78: 2707 (1981)

Weiss et al *J. Biomol Structure Dynamics* 1: 151 (1983)

2. 2D-NMR

Aue et al *J. Chem. Phys.* 64: 2229 (1976)

Wuthrich et al *J. Mol Biol.* 155: 311 (1982)

States et al *J. Mag. Res* 48: 286 (1982)

3. NMR of DNA

Scheck et al *JACS* 105: 2914 (1983)

Feigon et al *JACS* 104: 5540 (1982)

Hare et al *J Mol. Biol.* 71: 319 (1983)

Pardi et al *JACS* 105: 1652 (1983)

Weiss et al *Proc Natl Acad Sci USA* 81: 130 (1984)

Weiss et al *Nuc Acids Res* 12: 4035 (1984)