Multiple DNA-Protein Interactions Governing High-Precision DNA Transactions

HARRISON ECHOLS

The precise association of DNA-binding proteins with localized regions of DNA is crucial for regulated replication and expression of the genome. For certain DNA transactions, the requirement for precision in localization and control is extremely high. High-precision events amenable to detailed biochemical analysis are the initiation of DNA replication and site-specific recombination by bacteriophage λ and *Escherichia coli*. Recent experiments indicate that site-localization and control in these reactions involves the association of DNA-bound proteins to generate organized nucleoprotein structures in which the DNA is folded or wound. These specialized nucleoprotein structures are likely to provide the requisite accuracy for site localization and the necessary regulated reactivity to direct the DNA transaction. Multiple DNA-protein interactions are also required for controlled transcription of the eukaryotic genome. Distant upstream regulator and enhancer sequences may define protein-binding sites that form part of a reactive nucleoprotein structure capable of initiating transcription.

THE INITIATION OF DNA REPLICATION BY *Escherichia coli* occurs once per cell generation at a single site selected from the 4×10^6 DNA base pairs that constitute the genome (*1*). The recombinational insertion of the DNA of bacteriophage λ into the genome of *E. coli* takes place at a single site on the DNA of the phage and host; misinsertion of λ DNA at the aggregate of alternative sites in *E. coli* is less than 1/100 as frequent (*2*). The basis of the exceptional precision of these DNA transactions is not immediately evident from the DNA-binding properties of the proteins that localize these and other such reactions. Site-specific DNA-binding proteins have the general capacity to associate with DNA (nonspecific binding) and also typically bind tightly to improper sites that resemble closely (or are identical to) the "correct" binding sequence involved in control of the transaction (pseudosites) (*3*). Thus, high-precision interactions are unlikely to be localized by the binding of a single protein to a single DNA site.

All DNA transactions do not require identical precision. Prokaryotic transcription regulators must be highly accurate, but are likely to operate in a "moderate precision" mode because an error per cell generation is likely to be a tolerable number. However, to achieve the same biological accuracy the localization mechanism controlling eukaryotic transcription is likely to be much more stringent because some 10^{10} base pairs must be scanned by the regulators instead of the roughly 10^7 of their prokaryotic counterparts (*4*). Two troublesome pseudosites in the prokaryotic genome expand to 2000 in the eukaryotic. Thus, the biological accuracy problem is the ratio of sites scanned to sites selected. From this point of view, regulation of eukaryotic transcription can be considered a high-precision interaction.

In this article, I consider three high-precision interactions in *E. coli*: site-specific recombination by bacteriophage λ and the initiation of DNA replication by λ and by *E. coli*. The DNA-binding proteins that localize these reactions appear to use multiple DNA-protein and protein-protein interactions to generate a specialized nucleoprotein structure (snup) in which the DNA is folded or wound. I also briefly review the evidence that multiple DNA-protein interactions are responsible for the specificity of eukaryotic gene transcription. On the basis of prokaryotic work and additional inferences, snups seem likely to be responsible also for these high-precision reactions, as well as for site-specific recombination and initiation of DNA replication (*5, 6*).

Prokaryotic Transcription Regulators

Based on recent studies by x-ray crystallography, a simple and general picture has emerged for the interactions responsible for the stable location of specific sites by prokaryotic transcription regulators (*7–9*). The proteins utilize primarily a bihelical supersecondary structure to grip the B-form DNA helix; one helix penetrates the major groove of the DNA and reads the base code for the recognition site (*7–9*). The specificity of site recognition is enhanced by a dimeric protein recognizing a symmetric binding site. Noted initially for the λ Cro repressor protein, the bihelical interaction is supported by crystallographic studies of Cro (*10*), λ cI repressor (*11*), *E. coli* Crp activator (also termed CAP) (*12*), and *E. coli* Trp repressor (*13*). Homologies in amino acid sequence indicate that a number of other regulatory proteins have a similar recognition mode (*7, 8*).

The bihelical recognition mechanism is fundamentally an addition of the regulatory protein to the DNA helix without a drastic alteration in the linear duplex. The inferred mechanism of control is a protein-protein interaction in which the regulatory proteins either impede or facilitate productive binding by RNA polymerase (*7–9*). This simple addition interaction is probably sufficiently precise for most transcription regulation because an occasional error is unlikely to be serious. Although the addition interaction seems likely to be rather general, there are indications of greater effects on DNA structure, involving bending or looping, for the action of certain transcription regulators: the GalR repressor of the galactose operon (*14*) and the AraC activator of the arabinose operon (*15*), and even Crp (*16*). Thus, some instances of transcription regulation may be closer to the high fidelity interactions described below than is evident from the examples studied in detail so far.

The author is a professor in the Department of Molecular Biology, University of California, Berkeley, CA 94720.

Site-Specific Recombination by Phage λ

Phage λ integrates into and excises from the host *E. coli* genome by a regulated site-specific recombination event (*2, 17*). The recombination reaction requires three proteins. The λ Int protein is used for integration and excision, and the λ Xis protein for excision only; the integration host factor (IHF protein) is required for both directions (*2*). The overall recombination can therefore be written as:

$$\begin{array}{cccc} attP & attB & & attL \quad attR \\ POP' + BOB' & \overset{\text{Int/IHF}}{\underset{\text{Int/Xis/IHF}}{\rightleftharpoons}} & BOP' + POB' & \end{array}$$

In this formulation P, P′, B, and B′ refer to distinct segments of the phage (P) or host (B) DNA and O to the crossover region common to all sites. The phage (*att*P) and host (*att*B) sites recombine to yield the left (*att*L) and right (*att*R) sites of the inserted prophage. This section will review briefly the evidence that the exquisite precision and directional control of site-specific recombination involve specialized nucleoprotein structures. This point of view is developed more thoroughly in previous papers (*5, 18*).

Site-specific recombination has been analyzed at the DNA level by studies of Int-, Xis-, and IHF-binding sites and by determination of the extent of DNA required for the reaction (*2*). These studies have revealed a remarkable complexity in the phage *att*P site: some 230 bp (800 Å) are required, including seven binding sites for Int (two in P, two at O, and three in P′) (*2, 19–21*). The host *att*B site is simpler, involving about 20 bp and two Int sites (*2, 19–21*). The breaking and joining event, catalyzed by Int, occurs by a 7-base staggered cleavage in the crossover (O) region (*22, 23*). These studies pose an obvious question about mechanism: why devote 800 Å of DNA and seven Int-binding sites to a reaction that occurs in 24 Å of the crossover region?

Electron microscopic and topological studies provide strong evidence that the *att*P site does not enter the recombination reaction as the 230-bp linear DNA, but as part of a folded nucleoprotein structure. As viewed by the electron microscope, Int protein condenses the 800 Å of *att*P (POP′) DNA into a tight complex about 140 Å in diameter (*5, 18, 24*) (Fig. 1). Formation of this structure requires the P′ region because a localized (smaller) structure is found for *att*L, but not for *att*R or *att*B (*18*). Because Int appears to bind the P′ region with the highest affinity (*19, 25*), a plausible mechanism for formation of the specialized nucleoprotein structure is the sequential reaction shown in Fig. 2 (*24*); Int binds very tightly to the P′ region to initiate a series of protein-protein interactions with Int bound at O and P, generating the snup (Fig. 2, right). Because the Int-binding sites are not contiguous or in the same orientation, the substrate DNA must be folded or wound. Int-containing snups are termed "intasomes."

The form of the intasome active in recombination will also carry the IHF protein, which binds to three sites in *att*P DNA, and is required for site-specific recombination (*2, 26*). IHF and Int participate in a series of cooperative interactions for DNA binding, favored by supercoiling, that are also indicative of a higher order nucleoprotein structure (*27*). As judged by electron microscopy, the IHF protein is not required for intasome formation, although the structure may be more precise with IHF (*24*). Because a mutant Int protein (Inth) can carry out site-specific recombination without IHF, the IHF protein presumably has an ancillary (though crucial) role in normal integrative recombination (*28, 29*).

A second series of electron microscope experiments correlates snups more directly with recombinational reactivity. The Xis protein is required specifically for excisive recombination between the prophage *att*L and *att*R sites. Xis is also required for formation of a

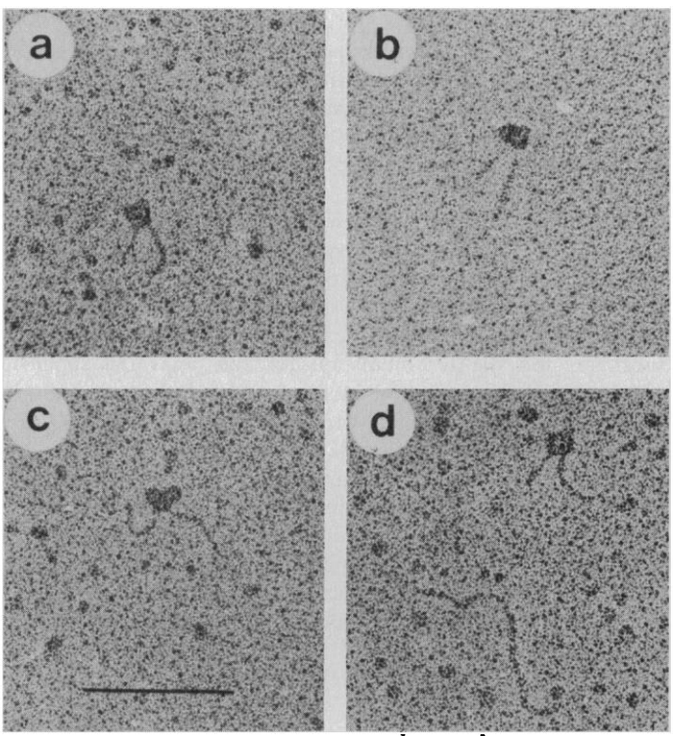

Fig. 1. Electron micrographs of the nucleoprotein structure formed by Int protein and the *att*P site. Int protein was bound to superhelical plasmid DNA carrying *att*P, and the plasmid DNA cleaved with restriction enzymes to produce the 493-bp *att*P fragment shown. The folding or winding of the *att*P DNA in the nucleoprotein structure can be seen visually in (d) by a comparison of the DNA lengths with Int bound (upper right) to Int not bound (lower left). Reproduced with permission from (*24*).

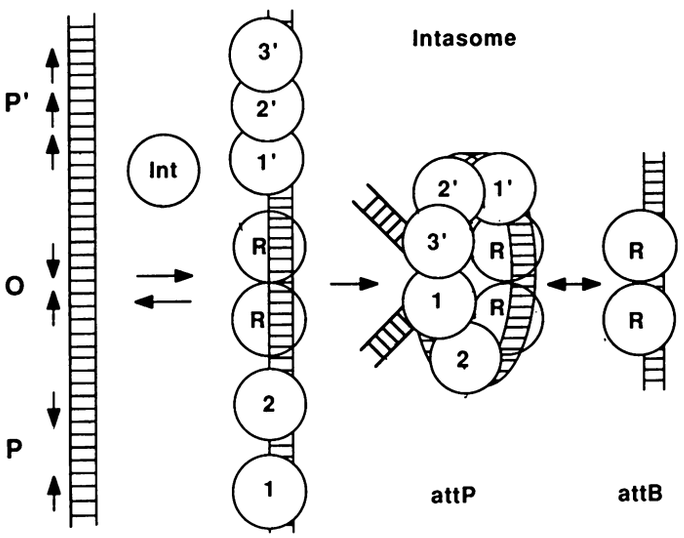

Fig. 2. Possible mechanism for formation of Int-mediated nucleoprotein structure (intasome). Int forms a high affinity complex in the P′ region (sites P1′, P2′, P3′) and lower affinity complexes at the recombining O region and at P (sites P1, P2). The Int molecules bound at P′ (designated 1′, 2′, 3′) nucleate a protein-protein association with Int bound in the O (R) and P (1, 2) regions to generate the DNA-wound nucleoprotein structure shown schematically on the right. The *att*P intasome is presumed to add two Int molecules bound at the O site of *att*B to produce specific pairing of the substrate sites in a reactive configuration for recombination; the breaking and joining will be catalyzed by the four Int molecules at the O region (designated R), two from *att*P and two from *att*B. In addition to the binding sites for Int, there are also sites for IHF and Xis.

localized nucleoprotein structure on *att*R DNA; Int alone does not suffice (*18*). The structures mediated by Int at *att*P and by Int and Xis at *att*R are specific nucleoprotein complexes. The DNA within each structure is highly condensed, and the region of DNA included is extremely precise for each substrate site. Thus, directional control of site-specific recombination appears to depend on formation of the appropriate snup.

The concept of specialized nucleoprotein structure is also supported by the topological properties of the recombination reaction. Int-mediated recombination is often carried out with a substrate carrying both *att*P and *att*B sites. The products are typically topologically complex because a supercoiled DNA is the preferred substrate, and the recombination reaction traps the supercoils present between the recombination sites (*30*). However, topological complexities persist, even if the reaction is carried out with a nonsupercoiled substrate (*31, 32*) or if the minimum complexity product is analyzed (*32*). For example, inversion of a nonsupercoiled substrate by integrative recombination produces trefoil knots with constant orientation (*33*). These topological properties of the recombination reaction strongly indicate that the substrate DNA is wrapped in a nucleoprotein structure, entrapping a supercoil (Fig. 2) (*31–33*). The topological properties of the Int-mediated reaction have been reviewed recently (*34*).

Clearly, formation of a complex nucleoprotein structure will localize the *att*P site with high precision. But the *att*B site must also be precisely localized, and precision in site recognition must be converted into reactivity for recombination. I surmise that these requirements are fulfilled by the formation of a nucleoprotein paired structure involving both *att*P and *att*B (*18, 24*) (Fig. 2). This structure can provide the initial highly specific recognition of recombining substrates and might in addition produce a structural change in the DNA duplex favoring formation of a base-paired synaptic structure (transition state complex). Similarly, excisive recombination can proceed through formation of a nucleoprotein paired structure of *att*L and *att*R (*18, 24*). Xis-dependent paired structures have been observed for *att*L and *att*R (*18*).

In summary, a considerable body of evidence supports the involvement of snups in determining the localization and directional control of site-specific recombination by bacteriophage λ. How general are such structures? Among other site-specific recombination mechanisms under study, multiple DNA-protein interactions are clearly involved for transposon Tn3 resolvase (Res) (*35*), and for the related Gin (*36*), Hin (*37*), and Cin (*38*) inversion systems. For Res, electron microscopic and topological evidence strongly supports the concept that a snup is involved in site-specific pairing and recombination (*39*). Although less defined biochemically so far, the

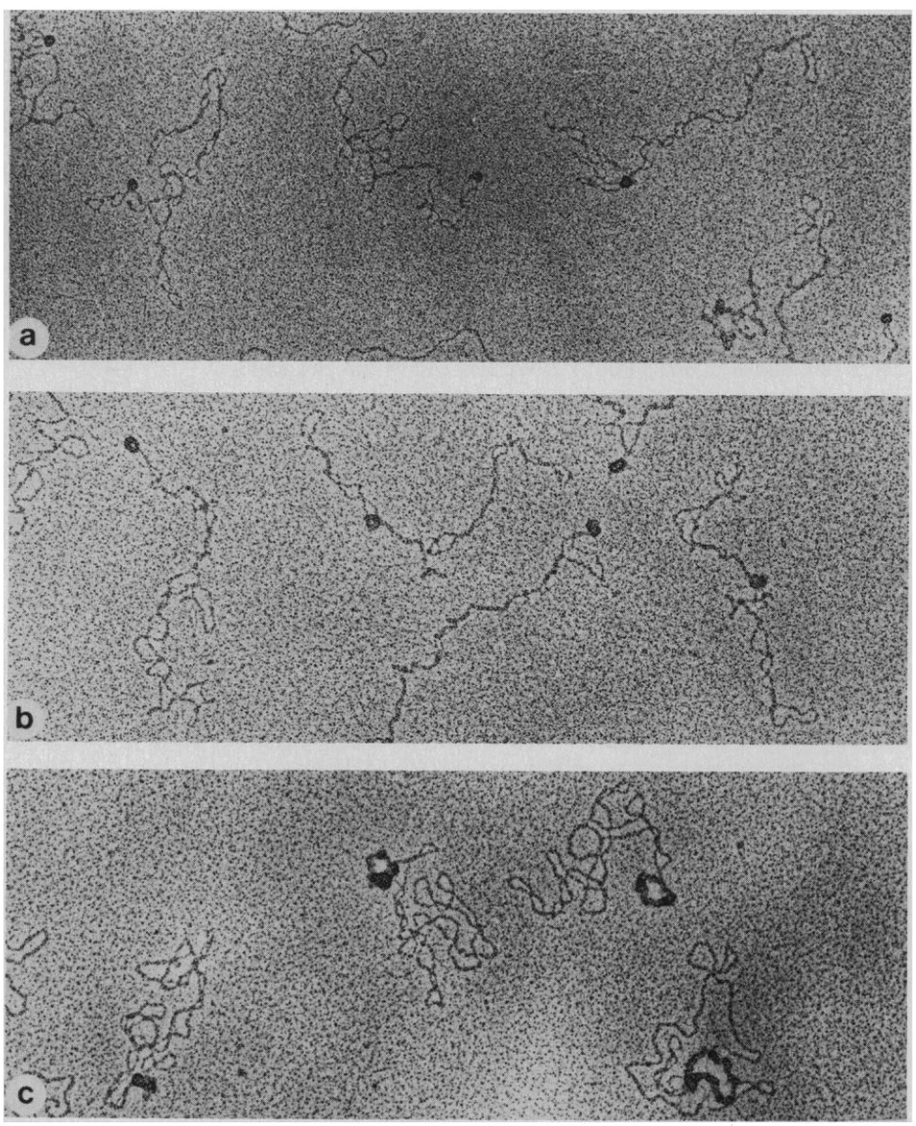

Fig. 3. Electron micrographs of nucleoprotein structures formed at different stages in the pathway to initiation of DNA replication at *ori*λ. (a) The O-mediated nucleoprotein structure presumed to localize precisely the replicational origin; (b) the augmented nucleoprotein structure formed by O, P, and DnaB; (c) the pre-initiation structure formed by adding the proteins DnaJ, DnaK, and Ssb, in which the replicational origin is locally unwound; (a) and (b) are reproduced with permission from (*48*); (c) was supplied by Mark Dodson (*49*).

Gin and Hin systems are intriguing because one critical binding site for an auxiliary protein is active in a number of locations and in either orientation, analogous to a eukaryotic enhancer sequence (*36, 37*). The likely role of snups in the precise initiation of DNA replication is considered next.

Initiation of DNA Replication by Phage λ and *E. coli*

The initiation of chromosomal DNA replication appears to proceed by a closely similar pathway for both phage λ and *E. coli*. For λ, the site-specific O protein recognizes the unique replication origin (*ori*λ), and with the λ P protein localizes critical *E. coli* replication proteins, including the DnaB helicase, the DnaG primase, and ultimately Pol III holoenzyme (*1, 40*). For *E. coli*, the site-specific DnaA protein recognizes the replication origin (*ori*C), and with the *E. coli* DnaC protein directs the addition of other replication proteins, including DnaB, DnaG, and eventually Pol III, to the localized start region for replication (*1*). Purified replication systems have been developed that utilize specifically the replication origins of *E. coli* (*41, 42*) and phage λ (*43*). The replication proteins can be used to analyze the mechanisms responsible for precise localization and reactivity in the initiation of DNA replication. The evidence that specialized nucleoprotein structures are used to initiate DNA replication is as follows.

The origin of λ DNA replication contains four 19-bp direct repeats, each of which is an inverted repeat (*44–46*). Purified O protein binds to the entire repeat region (*47*). Thus, the dimeric O protein probably has four binding sites, with each inverted repeat recognizing one dimer. Electron microscopic observations indicate that these multiple DNA-protein interactions lead to a specific nucleoprotein structure at the origin in which the DNA is folded or wound (*48*). The compact nucleoprotein structure (O-some) is presumably generated by protein-protein associations nucleated by the DNA-bound O molecules, possibly including additional O (*48–50*). The O-mediated nucleoprotein structure is likely to confer precise localization on the initiation site for replication.

This accurate recognition of the initiation site must be converted into specific reactivity. Recent experiments indicate that this occurs through a series of protein addition reactions, possibly accompanied by a structural change in the DNA duplex favoring origin-specific unwinding (*48, 49*). This series of events is shown in Figs. 3 and 4. The O-mediated nucleoprotein structure is shown in Fig. 3a. The addition of P and DnaB along with O results in a larger and more asymmetric structure at the replication origin (Fig. 3b). With the addition of DnaJ, DnaK, and the single-strand binding protein (Ssb), the origin DNA is unwound to yield single-stranded regions apparently coated by Ssb (Fig. 3c). The unwinding is probably catalyzed by the helicase activity of DnaB (*51*). The role of DnaJ and DnaK is probably to free DnaB from the multiprotein initiation complex (*49*). The locally unwound structure is presumably now competent to add the DnaG primase and the polymerase III replication enzyme. Thus, the O protein defines the replication origin and initiates a series of protein assembly events culminating in the localized initiation of DNA replication.

For *E. coli*, the origin of replication contains four noncontiguous 9-bp repeats in 240 bp of essential DNA (*52*). Purified DnaA protein binds to the 9-bp sequence either in the origin region or in other regions of DNA (*53*). However, only in the origin region does DnaA act to initiate DNA replication (*41, 42, 53*). The origin-specific interaction is characterized by a cooperative association with the entire 240-bp region, resulting in a nucleoprotein structure containing 20 to 30 DnaA molecules (*53*). Because the origin region has only four DnaA-binding sites, the initial interaction of DnaA with *ori*C DNA probably nucleates a protein-protein association to build the large multiprotein structure (analogous to assembly of a "mini" phage head). The DNA is likely to be folded or wound in the nucleoprotein complex, as judged by the large amount of DNA involved and the anomalous sensitivity of the DNA to nucleases (*53*). Thus, the DnaA protein appears to interact with its target site to generate a snup.

The fundamental mechanism by which DnaA localizes an active origin of replication is likely to parallel closely the sequence outlined above for O and *ori*λ, except that DnaC is used instead of λ P to bring in DnaB. Addition of DnaC and DnaB along with DnaA results in the formation of a larger structure (*54*). With the addition of DnaC, DnaB, Ssb, and DNA gyrase along with DnaA, a highly underwound DNA molecule is generated after protein removal (*42*).

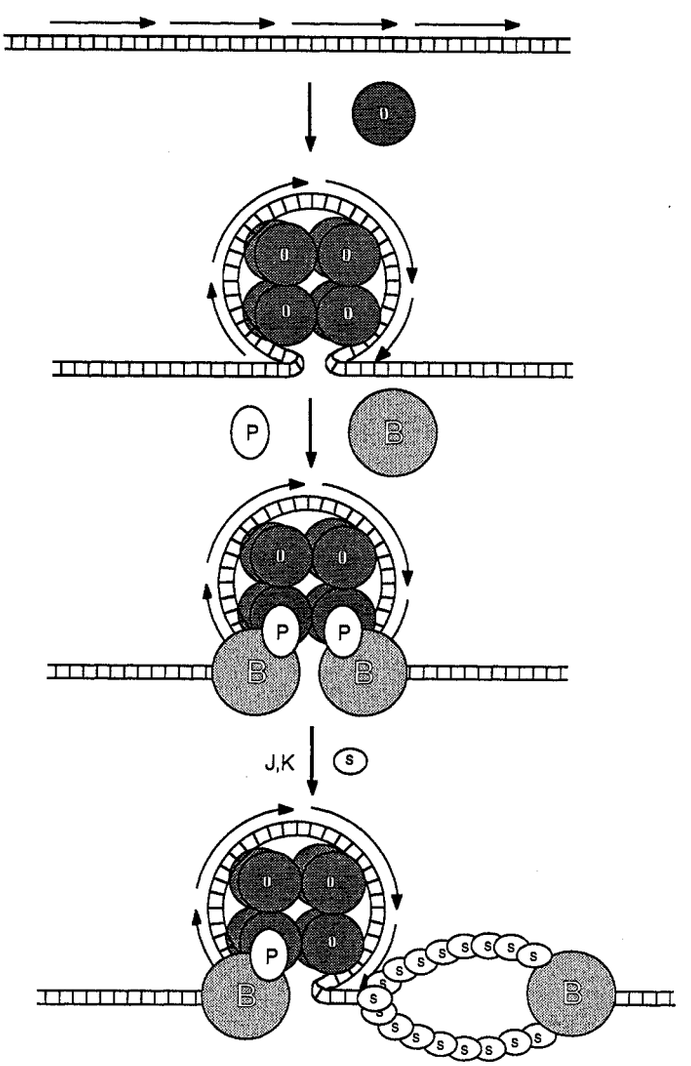

Fig. 4. Inferred pathway to initiation of DNA replication at *ori*λ. The O protein binds to the four direct repeats in *ori*λ and self-associates to form the "O-some" (possibly along with additional O molecules) (*48*). The P protein binds to O and to DnaB to generate a more complex nucleoprotein structure. The addition of DnaJ, DnaK, Ssb, and ATP allows DnaB to act as a helicase, unwinding the origin region. The locally unwound DNA, coated with Ssb, is presumed to serve as a substrate for DnaG primase to synthesize an RNA primer, as for single-stranded phage replication systems. DNA polymerase III holoenzyme can elongate this primer to initiate leading strand DNA replication and thereby start normal double-strand replication (*1*). The eventual route to bidirectional replication is not known; as studied so far, the localized unwinding reaction goes only in the direction shown (*49*).

This origin-specific structure presumably results from a DNA-mediated unwinding reaction (Fig. 4); the topoisomerase activity of gyrase facilitates an extensive reaction by removing the topological constraint of opposing superhelical turns, which otherwise limits the helicase reaction (*42*) [a similar highly unwound structure is found for the *ori*λ reaction if gyrase is added (*49*)]. In the absence of gyrase, a limited, origin-specific unwinding is generated, which can be visualized by electron microscopic analysis without deproteinization (*54*). There are differences between the *ori*λ and *ori*C reactions analyzed so far in vitro. The *ori*C reaction is stimulated by HU protein (a histone-like protein) and does not need DnaJ and DnaK (*42, 54*). Despite these differences, the overall similarities indicate a closely similar route to origin-specific initiation of DNA replication for λ and *E. coli* (*55*). The crucial step for reactivity at the origin appears to be the localization of DnaB in a state competent for helicase activity. To achieve this at a single site only once per replication cycle (cell generation for *E. coli*) is a most impressive accomplishment.

For the extensively studied λ and *E. coli* systems, there is strong evidence that snups locate precisely and provide for controlled activation of the sites used to initiate chromosomal DNA replication. How general is this snup mechanism? Although the data from other replication systems are less complete, multiple DNA-protein interactions are found frequently for the proteins that localize the origins of DNA replication. For prokaryotes, the π protein of plasmid R6K has seven binding sites (*56, 57*); the RepA protein of phage P1 has 14 sites, five of which are in the origin region (*58*); the E protein of plasmid F has at least nine sites, four of which are in the origin region (*59*) [the non-origin sites are likely to have a regulatory role (*58, 59*)]. The single-strand DNA phage φX174 is an interesting example of a different type, in which multiple DNA-protein interactions generate an appropriate single-strand DNA complex competent for priming (primosome) (*1*). For eukaryotes, the T antigen of SV40 virus has three binding sites (*60*); T antigen also associates with DNA polymerase α, indicating that a sequence of protein addition reactions may be involved in initiation of DNA replication by SV40 (*61*). The nuclear antigen (EBNA-1) protein of Epstein-Barr virus has at least 24 binding sites, four of which are in the origin region (*62*). Thus, the fundamental requirement for formation of snups, a multiple binding interaction, appears to be a widely found feature of replication origins.

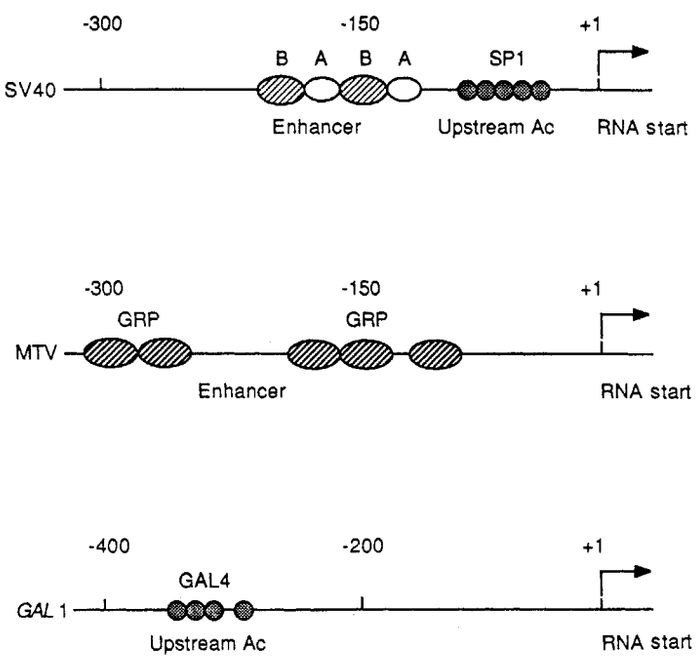

Fig. 5. Spatial relationship of activator sites and RNA start for some eukaryotic promoters. For the SV40 early transcript, the upstream activator region specifies five binding sites for protein SP1; the repeated enhancer sequence binds at least two proteins, designated A and B. The glucocorticoid receptor protein (GRP) activates transcription of integrated MTV; an upstream enhancer sequence specifies five binding sites for GRP. The GAL4 protein is a positive regulator for transcription of the *GAL1* gene of yeast; GAL4 binds to four sites in the upstream activator sequence. For all three transcription systems, the spatial relationship is inconsistent with a direct interaction on linear DNA between activator proteins and RNA polymerase at its start site.

Do Specialized Nucleoprotein Structures Control Eukaryotic Transcription?

The transcription of eukaryotic genes is clearly subject to regulation by DNA-binding proteins (*63–65*). Although this general feature is similar to prokaryotic regulation, the spatial properties of eukaryotic control sites are notably different. The regulatory sites are often a hundred or more base pairs from the start point of RNA synthesis, and certain regulatory sequences (enhancer sequences) can function in an orientation-independent fashion and even downstream from the gene that they regulate (*66*). The properties of these distant control sequences have led to the notion that the biochemical mechanism for regulation of eukaryotic genes might be fundamentally different from the direct protein-protein interaction that appears to prevail in prokaryotes. However, the existence of distant control sequences is also consistent with regulation by specialized nucleoprotein structures similar to those described above, in which DNA-bound proteins associate with each other to generate an active transcription complex at the start site for RNA synthesis (*6, 67, 68*). There is considerable evidence that multiple DNA-protein interactions are often involved in eukaryotic gene regulation, and it appears that these multiple binding interactions might act through specialized nucleoprotein structures.

The spatial differences between prokaryotic and eukaryotic regulation are most clearly evident in the case of the messenger RNA (mRNA) genes subject to polymerase II transcription. A few well-studied examples are shown diagrammatically in Fig. 5. Efficient transcription of the early promoter of SV40 virus requires two distinct control regions: the five upstream promoter sequences that are binding sites for the SP1 protein (*65, 69*), and the repeated enhancer region that probably binds at least two different proteins (A and B on Fig. 5) (*70, 71*). Transcriptional expression of an integrated mouse mammary tumor virus (MTV) is subject to regulation by glucocorticoid receptor protein (GRP); this protein binds to five sites upstream of the RNA start site, all of which appear to be required for maximal transcription (*72, 73*). The yeast genes for the proteins of galactose metabolism are controlled by the GAL4 regulatory protein; the GAL4 protein binds to the upstream activator region of the *GAL1* gene (*74, 75*), probably to four sites (*75*). In some way, the information for specificity of transcription must be transmitted from distant binding sequences to the promoter proximal sequences (for instance, the TATA box) to define the RNA start site (*76*).

Although the total number of examples is limited, four properties of upstream promoter sequences and enhancers seem likely to be widespread: (i) the control sequences are binding sites for regulatory proteins; (ii) the DNA-binding sites are multiple; (iii) the control sequences do not require a fixed distance from the RNA start site; and (iv) the distances between regulatory sites and RNA start site are often too large for a direct protein-protein interaction on linear DNA (*65, 77, 78*). These properties are consistent with

three general models for the control sequence: (i) an entry point for RNA polymerase or the transcription factor, from which it migrates to the RNA start site; (ii) a site for a DNA structural activator, from which a structural change is transmitted along the DNA (such as by a site-specific DNA topoisomerase); and (iii) a binding site for a protein that undergoes a DNA-bound interaction with RNA polymerase and other transcription factors (such as a specialized nucleoprotein structure).

At present, there are no definitive experiments that argue compellingly for any of the three models: migration, transmission, or protein association. The notion of a specialized nucleoprotein structure is appealing because an analogous interaction of spatially separated, DNA-bound proteins is involved in initiation of DNA replication and site-specific recombination in prokaryotes. Moreover, the snup structure provides a solution to the scanning problem noted in the introduction: the precise selection of a localized site from a very large spectrum of potential sites (*4, 6*). In addition to providing specificity, the use of highly multiple binding interactions that can act over a distance allows many possible (temporal or cell-specific) inputs into the decision whether to use a given promoter.

The limited experimental evidence for eukaryotic transcription appears to favor the nucleoprotein idea more than other possible mechanisms. For the transmission or migration models, a repressor is most likely to work either at the distant activator sequence or between the activator sequence and the RNA start. For the α2 repressor of yeast, the repressor normally binds to a sequence between the upstream activator and the RNA start (*79*). However, α2 will also bind and repress on the upstream side of the activator sequence (though less well) (*79*). Repression activity at a binding site upstream from the activator sequence is inconsistent with the transmission and migration models, but can be accommodated by the nucleoprotein mechanism with the assumption that the bound repressor interferes with formation of the activating snup.

Experiments with a positive regulatory protein of yeast, GAL4, are also indicative of protein-protein interactions. The GAL4 protein normally binds to the upstream activator sequence controlling the *GAL1* gene (Fig. 5) (*74, 75*). The DNA recognition region of the GAL4 protein can be replaced by the corresponding region of an *E. coli* protein, the LexA repressor, producing a hybrid activator protein with regulatory function dependent on the binding site for LexA (*67*). This hybrid protein activates transcription at two sites upstream from the transcription start and even works (though poorly) downstream from the gene it controls (*67*). Although compatible with some form of all three models noted above, these results are most consistent with the direct association of DNA-bound proteins.

For higher eukaryotes, an indirect inference for the nucleoprotein model stems from one property of the SV40 enhancer sequence: promoters differ in their protein requirements for enhancer activity (*68*). The SV40 enhancer can activate the SV40 early promoter and the herpes thymidine kinase promoter, but the activity for thymidine kinase also requires SV40 T antigen. One explanation for such diversity of activity exhibited by this and other enhancer systems is the formation of a multiprotein complex at the promoter involving enhancer-binding proteins, other specific activator proteins, and RNA polymerase (*68*).

In summary, current evidence appears to be most consistent with the concept that snups control transcription of eukaryotic mRNA genes. However, the direct demonstration of such structures with purified proteins and DNA is clearly required to verify these inferences. For the transfer RNA and 5*S* RNA genes transcribed by RNA polymerase III, a series of localized protein addition reactions has been defined in vitro for activating transcription (*63, 64*). The generation of a stable, multiprotein transcription complex is very similar to the snups described above for prokaryotes. The DNA-binding protein that localizes the RNA start site binds within the gene that it regulates, but close enough to the RNA start site for a direct protein-protein interaction with other transcription factors and RNA polymerase (*63, 64*). As judged by DNase protection and electron microscopy, the DNA-binding transcription factor for yeast transfer RNA genes may introduce a DNA-wrapping reminiscent of prokaryotic snups (*80*). Similar experiments with the proteins that bind to the distant activator sequences, the TATA region, and RNA polymerase itself should clarify the nature of the more complex interactions responsible for controlled transcription of mRNA genes.

REFERENCES AND NOTES

1. A. Kornberg, *Supplement to DNA Replication* (Freeman, San Francisco, 1982).
2. R. A. Weisberg and A. Landy, in *Lambda II*, R. W. Hendrix, J. W. Roberts, F. Stahl, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983), pp. 211–250.
3. P. H. von Hippel and O. G. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1608 (1986).
4. S.-Y. Lin and A. D. Riggs, *Cell* **4**, 107 (1975).
5. H. Echols, M. Better, M. Dodson, J. Roberts, R. McMacken, *Cold Spring Harbor Symp. Quant. Biol.* **49**, 727 (1984).
6. H. Echols, *BioEssays* **1**, 145 (1984).
7. Y. Takeda, D. H. Ohlendorf, W. F. Anderson, B. W. Matthews, *Science* **221**, 1020 (1984).
8. C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.* **53**, 293 (1984).
9. M. Ptashne, *A Genetic Switch, Gene Control and Phage λ* (Cell Press/Blackwell, Palo Alto, 1986).
10. W. F. Anderson, D. H. Ohlendorf, Y. Takeda, B. W. Matthews, *Nature (London)* **290**, 754 (1981).
11. C. O. Pabo and M. Lewis, *ibid.* **298**, 443 (1982).
12. I. T. Weber and T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3973 (1984).
13. R. W. Schevitz, Z. Otwinowski, A. Joachimiak, C. L. Lawson, P. B. Sigler, *Nature (London)* **317**, 782 (1985).
14. A. Majumdar and S. Adhya, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6100 (1984).
15. T. Dunn, S. Hahn, R. Schleif, *ibid.*, p. 5017.
16. A. Kolb, A. Spassky, C. Chapon, B. Blazy, H. Buc, *Nucleic Acids Res.* **11**, 7833 (1983); H. M. Wu and D. M. Crothers, *Nature (London)* **308**, 509 (1984).
17. H. Echols and G. Guarneros, in *Lambda II*, R. W. Hendrix, J. W. Roberts, F. Stahl, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983), pp. 75–92.
18. M. Better, S. Wickner, J. Auerbach, H. Echols, *Cell* **32**, 161 (1983).
19. P. W. Hsu, W. Ross, A. Landy, *Nature (London)* **285**, 85 (1980).
20. M. Mizuuchi and K. Mizuuchi, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3220 (1980).
21. W. Ross and A. Landy, *ibid.* **79**, 7724 (1982).
22. K. Mizuuchi *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **45**, 429 (1981).
23. N. L. Craig and H. A. Nash, *Cell* **35**, 795 (1983).
24. M. Better, C. Lu, R. C. Williams, H. Echols, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5837 (1982).
25. R. W. Davies, P. H. Schreier, M. L. Kotewicz, H. Echols, *Nucleic Acids Res.* **7**, 2225 (1979).
26. H. A. Nash and C. A. Robertson, *J. Biol. Chem.* **256**, 9246 (1981); N. L. Craig and H. A. Nash (*23*).
27. E. Richet, P. Abcarian, H. A. Nash, J. F. Thompson, L. Moitoso de Vargas, S. E. Skinner, personal communication; A. Landy, personal communication.
28. H. I. Miller, M. M. Mozola, D. I. Friedman, *Cell* **20**, 721 (1980).
29. B. J. Lange-Gustafson and H. A. Nash, *J. Biol. Chem.* **259**, 12724 (1984).
30. K. Mizuuchi, M. Gellert, R. Weisberg, H. Nash, *J. Mol. Biol.* **141**, 485 (1980).
31. T. J. Pollock and H. A. Nash, *ibid.* **170**, 1 (1983).
32. S. J. Spengler, A. Stasiak, N. R. Cozzarelli, *Cell* **42**, 325 (1985).
33. J. D. Griffith and H. A. Nash, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3124 (1985).
34. S. Wasserman and N. Cozzarelli, *Science* **232**, 951 (1986).
35. N. D. F. Grindley, M. R. Lauth, R. G. Wells, R. J. Wityk, J. J. Salvo, *Cell* **30**, 19 (1982).
36. R. Kahmann, F. Rudt, C. Koch, G. Mertens, *ibid.* **41**, 771 (1985).
37. R. C. Johnson and M. I. Simon, *ibid.*, p. 781.
38. H. E. Huber, S. Iida, W. Arber, T. A. Bickle, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3776 (1985).
39. H. Benjamin and N. Cozzarelli, personal communication.
40. M. E. Furth and S. H. Wickner, in *Lambda II*, R. Hendrix, J. W. Roberts, F. Stahl, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983), pp. 145–173.
41. A. Van der Ende, T. A. Baker, T. Ogawa, A. Kornberg, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3954 (1985).
42. T. A. Baker, K. Sekimizu, B. E. Funnell, A. Kornberg, *Cell* **45**, 53 (1984).
43. R. McMacken, K. Mensa-Wilmot, C. Alfano, M. Wold, personal communication.
44. G. Scherer, *Nucleic Acids Res.* **5**, 3141 (1978).
45. R. Grosschedl and G. Hobom, *Nature (London)* **277**, 621 (1979).
46. D. Moore *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **43**, 155 (1981).
47. T. Tsurimoto and K. Matsubara, *Nucleic Acids Res.* **9**, 1789 (1981); K. Zahn and F. Blattner, *EMBO J.* **4**, 3605 (1985).
48. M. Dodson, J. Roberts, R. McMacken, H. Echols, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4678 (1985).
49. M. Dodson *et al.*, *ibid.*, in press.
50. An O-some can be formed with as few as two of the direct repeats; this observation

indicates that the DNA-bound O may nucleate a "self-assembly" reaction of O dimers analogous to phage head assembly (as for the DnaA case described below). There is no evidence available as to the relative stability of the four-repeat O-some and the two-repeat structure.
51. J. H. LeBowitz and R. McMacken, *J. Biol. Chem.* **261**, 4738 (1986).
52. J. W. Zyskind, J. M. Cleary, W. S. A. Brusilow, N. E. Harding, D. W. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1164 (1983).
53. R. S. Fuller, B. E. Funnell, A. Kornberg, *Cell* **38**, 889 (1984).
54. T. A. Baker, B. E. Funnell, A. Kornberg, personal communication.
55. Additional complexities remain. The *ori*λ initiation reaction in vivo is bidirectional and strongly stimulated by RNA polymerase (*40*); in vitro the unwinding reaction is unidirectional and does not require RNA polymerase. The *ori*C initiation reaction in vivo is bidirectional and probably depends on RNA polymerase (*1*). The unwinding reaction in vitro is bidirectional; RNA polymerase is not required but is important under appropriate conditions (*42, 54*).
56. J. Germino and D. Bastia, *Cell* **34**, 125 (1983).
57. M. Filutowicz, E. Uhlenhopp, D. R. Helinski, *J. Mol. Biol.* **187**, 225 (1986).
58. A. L. Abeles, K. M. Snyder, D. K. Chattoraj, *ibid.* **173**, 307 (1984) and personal communication.
59. T. Tokino, T. Murotsu, K. Matsubara, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
60. R. Tjian, *Cell* **13**, 165 (1978).
61. S. T. Smale and R. Tjian, personal communication.
62. D. R. Rawlins, G. Milman, S. D. Hayward, G. S. Hayward, *Cell* **42**, 859 (1985).
63. A. B. Lassar, P. L. Martin, R. G. Roeder, *Science* **222**, 740 (1983).
64. D. D. Brown, *Cell* **37**, 359 (1984).
65. W. S. Dynan and R. Tjian, *Nature (London)* **316**, 774 (1985).
66. H. R. Scholer and P. Gruss, *Cell* **36**, 403 (1984).
67. R. Brent and M. Ptashne, *ibid.* **43**, 729 (1985).
68. P. Robbins, D. Rio, M. Botchan, *Mol. Cell. Biol.* **6**, 1283 (1986).
69. W. S. Dynan and R. Tjian, *Cell* **35**, 79 (1983).
70. P. Sassone-Corsi, A. Wildeman, P. Chambon, *Nature (London)* **313**, 438 (1985).
71. A. G. Wildeman *et al.*, *Mol. Cell. Biol.* **6**, 2098 (1986).
72. F. Payvar *et al.*, *Cell* **35**, 381 (1983).
73. D. DeFranco, O. Wrange, J. Merryweather, K. R. Yamamoto, in *Genome Rearrangement*, I. Herskowitz and M. Simon, Eds. (Liss, New York, 1985), p. 305.
74. R. Bram and R. Kornberg, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 43 (1985).
75. E. Giniger, S. M. Varnum, M. Ptashne, *Cell* **40**, 767 (1985).
76. Of the examples in Fig. 5, only for SP1 has a transcription activation function been demonstrated in vitro; this limitation is not surprising given the complexity of the available transcription systems, but it serves to emphasize the preliminary nature of the conclusions that can be drawn.
77. L. Guarente, *Cell* **36**, 799 (1984).
78. S. McKnight and R. Tjian, *ibid.*, in press.
79. A. D. Johnson and I. Herskowitz, *ibid.* **42**, 237 (1985).
80. D. J. Stillman, M. Better, E. P. Geiduschek, *J. Mol. Biol.* **185**, 451 (1985).
81. I thank Michael Botchan, Nicholas Cozzarelli, Arthur Kornberg, Roger McMacken, and Robert Tjian for valuable discussion and critical reading of the manuscript. I thank Marc Better and Mark Dodson for their many contributions to the ideas discussed here, Terri DeLuca for editorial help, and Mark Dodson and Chris Vasilikiotis for figures. Research on this topic in my laboratory has been supported by a grant from the National of Institutes of Health (GM17078).

Natural Philosophy in the Constitution

GERARD PIEL

The natural philosophers who wrote the U.S. social contract held the advancement of science to be the supreme exercise of citizen sovereignty. The rising nation, in the late 19th century, established the seat of that sovereignty in its universities. Today those institutions have come to be regarded as contract research centers at the service of the federal government. Research contracts in support of the proposed Strategic Defense Initiative are pressed on them against the consensus of the scientific community that holds this "Star Wars" enterprise to be technically infeasible. The time has come to reconstruct the relation between the federal government and university science in the spirit of our social contract.

AS THE SOCIOLOGY OF SCIENCE HAS SHOWN US, THE seeking of consensus is the habit that gets the work of science done. This social process goes forward simultaneously and consecutively in two phases, private and public. In the image of Galileo standing before the most awesome power in his world and time, the scientist can accept no authority but his own lonely reason and judgment. Yet, in the words of Robert K. Merton, founder of the sociology of science, it is "only after the originality and consequence of [his] work have been attested by significant others [the colleagues most closely engaged in his work] can the scientist feel reasonably comfortable about it" (*1*). Those others arrive at consensus not by taking a vote but by the same lonely exercise of reason and judgment.

Scientists tend to carry this habit over into their consideration of public issues. As Jerome Wiesner has observed, "Reasonable men in possession of the same set of facts tend to arrive at the same conclusion" (*2*). Issuing from perhaps the only community in society capable of forming assured consensus, the consensus of the scientific community on public issues ought to be more widely recognized in the deliberations of our federal government.

Consensus Unheard

For reasons that betray ignorance of science among persons who have a responsibility to know better, however, that consensus goes largely unheard. The title of scientist is clothed with received authority in its most antiscientific mode. The title is, moreover, indiscriminately bestowed. In accordance with custom—and from failure to exercise professional judgment—the press almost invariably gives equal time to the consensus of the community, on the one hand, and to the eccentric celebrity and others equally unqualified to speak about the topic at issue, on the other hand (*3*).

Among the citizenry at large there is little understanding of the social process of science. Worse yet, there is widespread misunder-

The author is the chairman of the board, *Scientific American*, New York 10017, and the retiring president of AAAS. This article is based on his lecture at the AAAS annual meeting on 28 May 1986 in Philadelphia.