Pure & Appl. Chem., Vol. 57, No. 1, pp. 45–56, 1985.
Printed in Great Britain.

Thermodynamic aspects of the regulation of protein synthesis in bacteria[1]

Peter H. von Hippel and Frederic R. Fairfield

Institute of Molecular Biology and Department of Chemistry
University of Oregon
Eugene, Oregon 97403 USA

Abstract - Gene expression, manifested as the orderly production of specific proteins of appropriate types and amounts in defined temporal progressions, is regulated at virtually every step of messenger RNA transcription and its subsequent translation into protein. The mechanisms whereby these processes are controlled are all physical chemical in nature; i.e., their functions must be explicable in terms of sets of coupled equilibrium and/or kinetic parameters.

In this article we discuss the physical chemical bases of two such control systems that seem to be straightforwardly thermodynamic in nature: (i) the autoregulation, at the translational level, of the production of a single-stranded DNA binding protein essential for DNA replication; and (ii) the coordinate control of the production of the proteins required for the self-assembly of the bacterial ribosome. Quantitative models for both systems are presented, and the thermodynamic principles underlying such models are discussed in general terms.

INTRODUCTION

Cell function and development require the orderly synthesis and utilization of a myriad of different proteins. Some of these proteins function as single subunit species in solution, but most are assembled, transiently or permanently, into structures of considerable complexity (e.g. replication or transcription complexes, ribosomes, mitotic constituents, etc.). The central regulatory problem for cells is to control the amounts and types of proteins that are synthesized in response to the demands of the cell cycle, of developmental processes, and of changing external conditions.

Since proteins are the end-product of gene expression, their synthesis can be regulated at the level of gene duplication (DNA production), transcription (messenger RNA production) or translation (protein production). All these modes of regulation have advantages and disadvantages. Gene duplication is indirect, and somewhat inefficient, but can result in large amplification of protein production (for a recent review see Ref. 1). Transcriptional control is more flexible and efficient, and a variety of elegant molecular mechanisms have evolved that regulate mRNA synthesis at virtually every step of its production (see Ref. 2). However control of protein synthesis via regulation of mRNA production is still indirect, and thus, depending on circumstances, can be slow and quantitatively and temporally imprecise.

It is becoming increasingly clear that regulation directly at the level of protein synthesis (translation) has advantages, especially for genes that must produce large quantities of protein in defined bursts in response to cellular demand, with intervening periods in which no synthesis is required. There are a number of ways in which translation can be regulated (see Ref. 3); the simplest method probably involves a direct autogenous repression mechanism in which the protein (gene product) itself reversibly binds to a translational initiation (ribosome binding) site on its own mRNA and thus selectively (and reversibly) shuts off its own synthesis.

In principle a simple titration of the mRNA binding site should suffice in such regulatory systems. Thus, the concentration (activity) of the free protein in the cell rises until it approaches the reciprocal of its binding affinity for the control site on the mRNA. The control site is then titrated and translation of that mRNA species ceases until the concentration of free protein again drops in the cell, bringing about dissociation of the protein from the mRNA binding site and triggering further synthesis. Systems that work on

[1]Portions of this article have been adapted and updated from an earlier review on the same subject (Ref. 27).

this general principle have been elucidated by Gold and coworkers (Ref. 4) in describing the control of the synthesis of the single-stranded DNA binding protein coded by bacteriophage T4 (gene 32 protein), and by Nomura and coworkers (Ref. 5) in studying the regulation of synthesis of the ribosomal proteins of E. coli.

Three problems arise in systems of this sort, and all must be considered quantitatively at the physical chemical level in order to achieve a description of an actual functional control system.

(i) Many proteins that are translationally regulated do not function in isolation, but are assembled into complexes. Thus the "effective" binding constant involved in assembling the gene product onto its functional target must exceed the "regulatory" binding constant to the mRNA site. This permits assembly to go to completion before the "shut-off" concentration is achieved in solution and the mRNA site is titrated. Assuming that the entire system works on an equilibrium basis, this requires consideration of all the coupled equilibria of the cell in which the subject protein participates.

(ii) A simple titration is, by definition, "broad" on a free ligand (protein) concentration axis. To move from 50% titration to 90% titration of a simple binding site requires a factor of 10 increase in free protein concentration. To titrate the next 9% of the sites (to a total of 99%) requires another factor of 10 increase in the free protein concentration. For proteins needed by the cell in large quantities (e.g., single-stranded DNA binding proteins or ribosomal proteins) utilization of the protein and/or its shut-off must be effectively cooperative in the free concentration(s) of the protein(s) synthesized, in order to avoid the need for vast overproduction of protein for purely regulatory purposes.

(iii) Finally, in order to permit simple operation of the system and to optimize protein design, it is helpful if the same binding interaction of the protein is utilized in both regulatory and functional interactions.

To apply these considerations to the analyis of a given regulatory system, a number of binding and cooperativity parameters must be measured or inferred. This has been done for

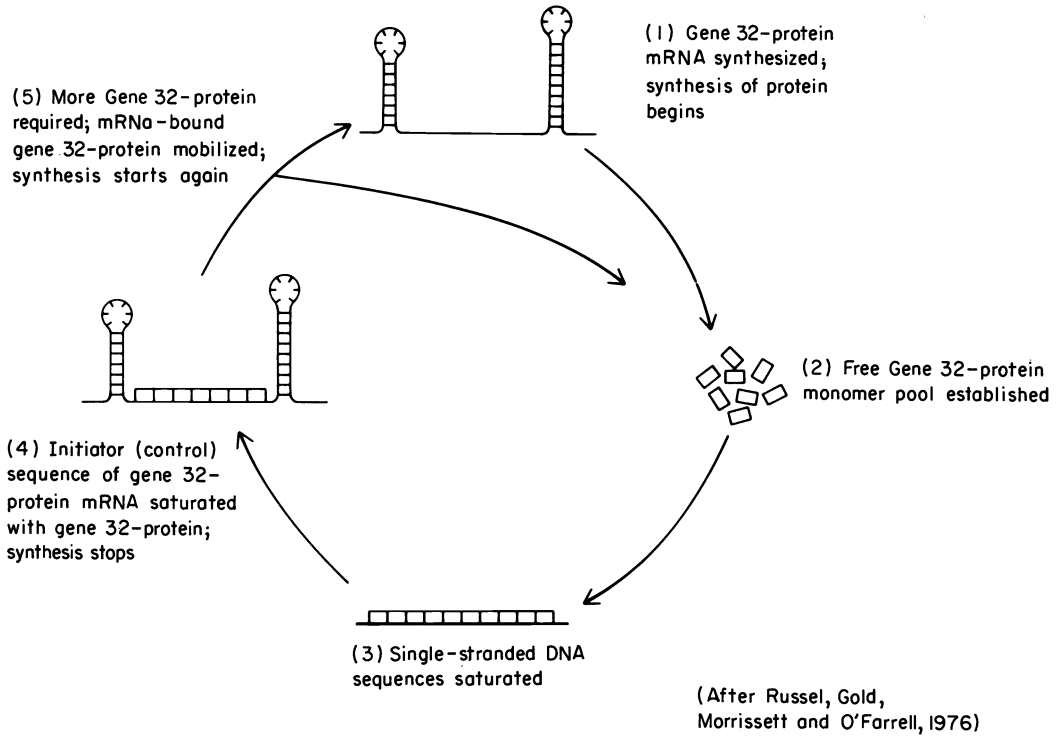

Fig. 1. Schematic representation of the sequence of events involved in the autogenous regulation of gene 32 protein synthesis, as elucidated by Gold and coworkers (Ref. 9,10). The various regulatory binding equilibria involved are indicated (see text).

the bacteriophage T4-coded gene 32 protein, and a complete analysis of this relatively simple autogenous regulatory system has been presented (Ref. 6). In this article the analysis of the gene 32 system is summarized, and then we consider the application of the same principles to the much more complex problem of the regulation of the synthesis of the ribosomal proteins of E. coli.

AUTOREGULATION OF BACTERIOPHAGE T4-CODED GENE 32 PROTEIN SYNTHESIS

Gene 32 protein is an essential component of the T4 DNA replication, recombination and repair systems (Ref. 8). It plays a "structural" (as opposed to a catalytic) role, binding in saturating amounts to the single-stranded DNA that is transiently produced in the essential intermediate stages of these processes. Genetic and biochemical studies have shown that the total amount of gene 32 protein produced in a phage infection depends directly on the amount of intracellular single-stranded DNA present (Ref. 7,8). It has also been shown that the synthesis of gene 32 protein is regulated at the translational level (Ref. 9,10).

In effect, intracellular control of the free concentration of gene 32 protein involves an orderly progression of binding events. All single-stranded DNA sequences are saturated as the level of free protein increases initially. Only after this process is complete does the free intracellular protein concentration rise to a level high enough to permit binding to the gene 32 mRNA "translational operator" site (Ref. 9,11), resulting in the specific cessation ("repression") of gene 32 protein synthesis. In vitro experiments have shown that this level of free protein concentration is not sufficient to permit binding to translational initiation sites of other T4 mRNAs (Ref. 10), to permit binding to the very large reservoir of double-stranded DNA present in the cell (Ref. 12,13), or to prevent the reannealing of double-stranded DNA after the replication process is complete (Ref. 6). This gene 32 protein functional and regulatory cycle is shown schematically in Fig. 1.

A combination of biochemical (Ref. 10) and physical chemical (Ref. 12,13,14,15) experiments has provided the necessary data for a quantitative molecular description of this autoregulatory cycle, which is responsible for the establishment and maintenance of physiological levels of gene 32 protein in T4 infection of E. coli. These studies are summarized briefly below; for further details see (Ref. 6), and citations therein.

Binding Parameters for Gene 32 Protein. The binding of a protein to a nucleic acid lattice can be described by three thermodynamic constants (Ref. 16). These are: (i) the binding site size (n; in units of nucleotide residues covered per protein monomer bound); (ii) the intrinsic association constant (K; in units of M^{-1}), and the cooperativity parameter (ω; unitless). These parameters have been measured for the binding of gene 32 protein to a variety of single-stranded deoxyribose- and ribose-containing homo- and heteropolynucleotides as a function of salt concentration and temperature (Ref. 12,13,14,15). The results show that n is constant at 7 ($\pm$1) nucleotide residues, that ω is constant at $\sim 2 \times 10^3$, and that K varies with nucleotide composition of the lattice, salt concentration and temperature (Ref. 13,15). These measurements have permitted us to calculate values of the effective affinity constant for gene 32 protein binding in the cooperative polynucleotide binding mode ($K\omega$) to single-stranded DNA and RNA sequences of T4 DNA of either known sequence or of average composition (Ref. 6), under physiological conditions defined as a temperature of 37°C, and a salt concentration of 0.23 M NaCl. (These conditions are equivalent, in terms of salt effects on the strength of protein-nucleic acid binding interactions, to the actual intracellular ionic environment of E. coli; see Ref. 17).

In Vitro Repression Experiments. Lemaire et al. (Ref. 10) have conducted experiments that demonstrate the translational repression of gene 32 protein in vitro, using a cell-free translation system containing a crude RNA preparation from T4-infected E. coli cells, and ribosomes, tRNA and supernatant proteins derived from uninfected E. coli. The results of these and other experiments may be summarized as follows: (i) Gene 32 protein binds preferentially to a specific component of the RNA derived from T4-infected cells. Since shut-off is specific for the synthesis of gene 32 protein, this component must be a portion of the gene 32 mRNA. (ii) The abruptness with which shut-off occurs as a function of added gene 32 protein suggests that this repression (and the binding of the protein to the gene 32 mRNA that is assumed to be responsible for it) must be cooperative in gene 32 protein concentration. (iii) Single-stranded DNA effectively binds gene 32 protein more tightly than single-stranded RNA, and, in particular, more tightly than the gene 32 mRNA "translational operator" (ribosome binding) site. (iv) The binding affinity of gene 32 protein for the gene 32 mRNA operator is larger than that for most other RNA constituents in the system, and is comparable to that of (unstructured) poly rU. (v) Double-stranded DNA, and also the other components of the cell-free translation system, bind gene 32 protein less strongly than does the gene 32 mRNA operator. (vi) The addition of gene 32 protein to levels that are 3- to 4-fold greater than required to halt gene 32 protein synthesis does shut off the synthesis of other T4 proteins in the cell-free translation system, suggesting that the gene 32 mRNA operator site differs only quantitatively (in

terms of gene 32 protein binding) from translational control sites on other T4 mRNAs. These and other data can also be used to estimate that the free intracellular gene 32 protein concentration maintained in vivo (during T4 infection) is ~3μM (Ref. 6).

Calculation of In Vivo Gene 32 Protein Binding (Titration) Curves for Various Structured and Unstructured Nucleic Acid Targets. Using the known binding parameters for gene 32 protein to various nucleic acid sequences, titration curves for the binding of gene 32 protein to various potential nucleic acid targets under physiological conditions have been calculated (Ref. 6). The results are fully and quantitatively compatible with the experimental facts outlined above, and together with the sequencing data of Krisch and coworkers (Ref. 7,18), have permitted the definition of the gene 32 mRNA translational operator site (Ref. 6).

A two-state calculation was used initially to determine the expected levels of binding of gene 32 protein to unstructured single-stranded DNA and RNA lattices. The results showed that long single-stranded DNA lattices of average T4 composition, unencumbered by secondary structure, would be expected to saturate at ~0.01 μM free gene 32 protein, while comparable RNA lattices would saturate at ~0.3 μM protein (Ref. 6). Both types of lattices should thus be fully saturated at physiological gene 32 protein concentrations.

However, most nucleic acid sequences in the cell are partially or completely involved in secondary structure. As a consequence the favorable binding free energy change (ΔG°_{bind}) involved in the interaction of gene 32 protein with single-stranded lattices will be opposed by the conformational free energy (ΔG°_{conf}) favoring the maintenance of partially double-stranded structures. This conformational free energy can be estimated using the approach and parameters developed by Tinoco and co-workers (Ref. 19). As a consequence higher free gene 32 protein concentrations are needed to saturate such initially structured nucleic acid lattices. Such calculations (not shown) reveal that, because of its tighter binding to single-stranded DNA lattices, physiological concentrations of gene 32 protein will saturate DNA lattices containing "stem-loop" structures with as much as 70% of the sequences involved in base-pairing. Thus virtually all secondary structure that might be expected to develop adventitiously in single-stranded regions during DNA replication should be removable by gene 32 protein at the controlled in vivo free protein concentration.

The situation for mRNA should be quite different. A variety of lines of evidence [see von Hippel et al. (Ref. 6) for a summary] suggest that mRNA (and ribosomal RNA; see below) secondary structure is crucial for biological activity, and thus should not be "melted" by gene 32 protein. The calculated results (Fig. 2) are fully compatible with this expectation, showing that because of the lesser affinity of gene 32 protein for

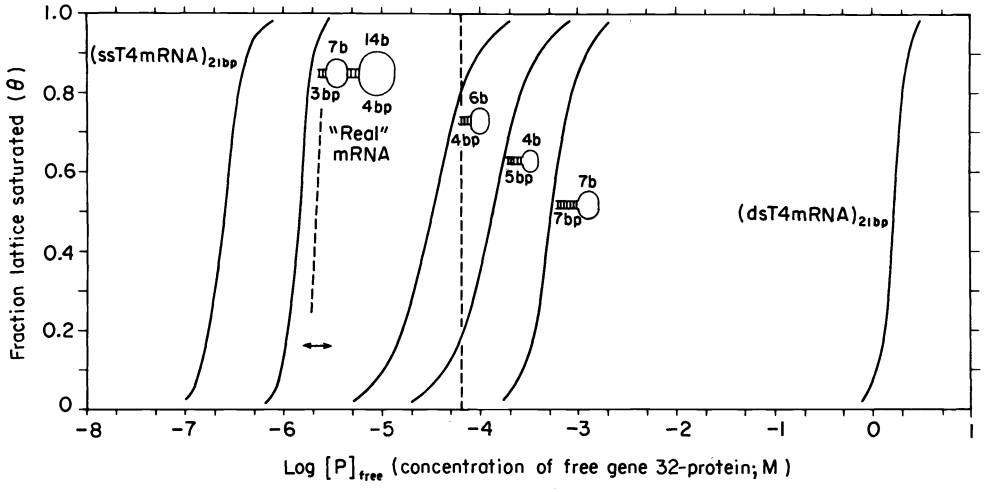

Fig. 2. Binding curves for the "melting" and complexation by gene 32 protein of various hypothetical initially looped and bulged T4 mRNA structures, plotted as a function of free gene 32 protein. The titration curves correspond, respectively, to the indicated stem-loop (and/or bulge) structures. The sloped dashed line labelled "Real mRNA" is the approximate binding isotherm for the gene 32 mRNA control site, as estimated from the Lemaire et al. experiments (Ref. 10). [Figure from Ref. 6.]

single-stranded RNA (relative to single-stranded DNA), only very "weak" elements of mRNA secondary structure should be "melted" at physiological gene 32 protein concentrations.

Finite Nucleic Acid Lattice Effects. To this point the calculations described above were carried out using a two-state "infinite lattice" model. In this model, it is assumed that the (e.g.) stem-loop regions for which binding curves are being calculated are already flanked by sites complexed with gene 32 protein. This means that every bound protein monomer will contribute a full "unit" of both intrinsic binding affinity (K) and binding cooperativity (ω) to the interaction. Thus:

$$\Theta = \frac{(K_{conf})(K_{bind})[P]^m}{1 + (K_{conf})(K_{bind})[P]^m} \quad (1)$$

where Θ = the fraction of the lattice sites under consideration that have been saturated at free protein concentration [P], m = the length of the lattice sequence under consideration in protein monomer units (m = N/n; where n = the protein site size and N = the lattice segment length in nucleotide residues), $K_{conf} = [NA_{ss}]/[NA_{ds}]$ (and $[NA_{ss}]$ and $[NA_{ds}]$ represent, respectively, the molar concentrations of open (single-stranded) and duplex (base-paired) nucleic acid lattice, in units of nucleotide residues), and:

$$K_{bind} = (K\omega)_1 (K\omega)_2 \cdots\cdots (K\omega)_m = \prod_{i=1}^{i=m} (K\omega)_i \quad (2)$$

We note that $K_{bind} = (K\omega)^m$ for infinite lattices of constant composition.

This model is quite appropriate for considering the titration, by gene 32 protein, of an mRNA segment containing a "weak" stem-loop structure ("hairpin"), or for "filling in" a single-stranded lattice segment comprising the moving "single-stranded window" in a moving DNA replication fork, but is less valid for considering the saturation of single-stranded regions within an mRNA molecule that are flanked by elements of secondary structure too stable to be "melted" at the physiological gene 32 protein concentration. For such regions a finite lattice calculation needs to be made, where:

$$K_{bind} = K_1 (K\omega)_2 (K\omega)_3 \cdots (K\omega)_m = K_1 \prod_{i=2}^{i=m} (K\omega)_i \quad (3)$$

We note that the finite lattice binding definition of K_{bind} (eq. 3) differs from that for infinite lattice binding (eq. 2) only by the loss of one "unit of ω", but, as Fig. 3 shows, for short sequences this loss can make an enormous difference in the resulting titration curve. (See Ref. 6 for further details.) Therefore, due to this "finite lattice" effect,

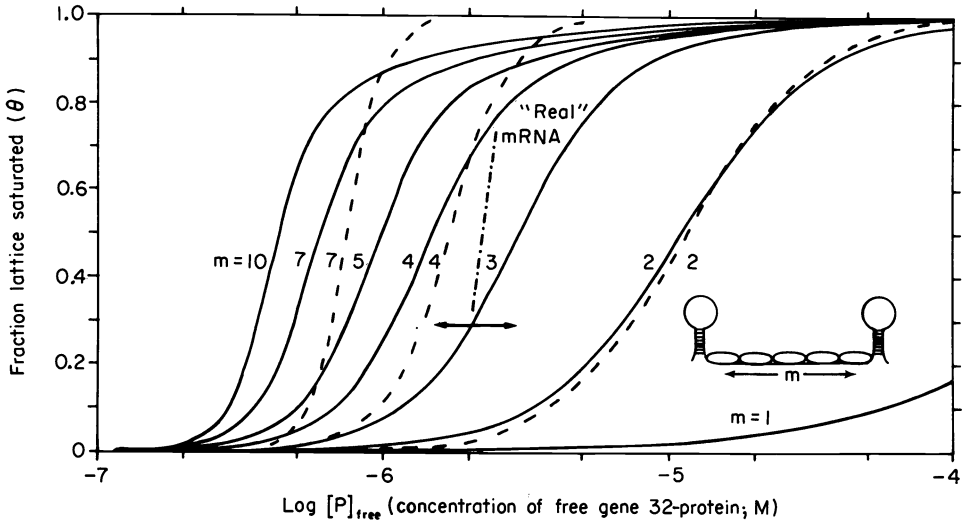

Fig. 3. Binding curves for the finite mRNA lattices of varying length. The dashed curves represent the two state approximation, calculated as outlined in the text. The solid curves were calculated by the "exact" method of Epstein; for further details see Newport et al. (Ref. 13). The lengths of the lattices are defined in units (m) of protein monomer binding sites. The site size of gene 32 protein binding cooperatively in the polynucleotide binding mode is 7 nucleotide residues. Thus the lengths of the respective finite lattices, in units of nucleotide residues, are 7m). [Figure from Ref. 6.]

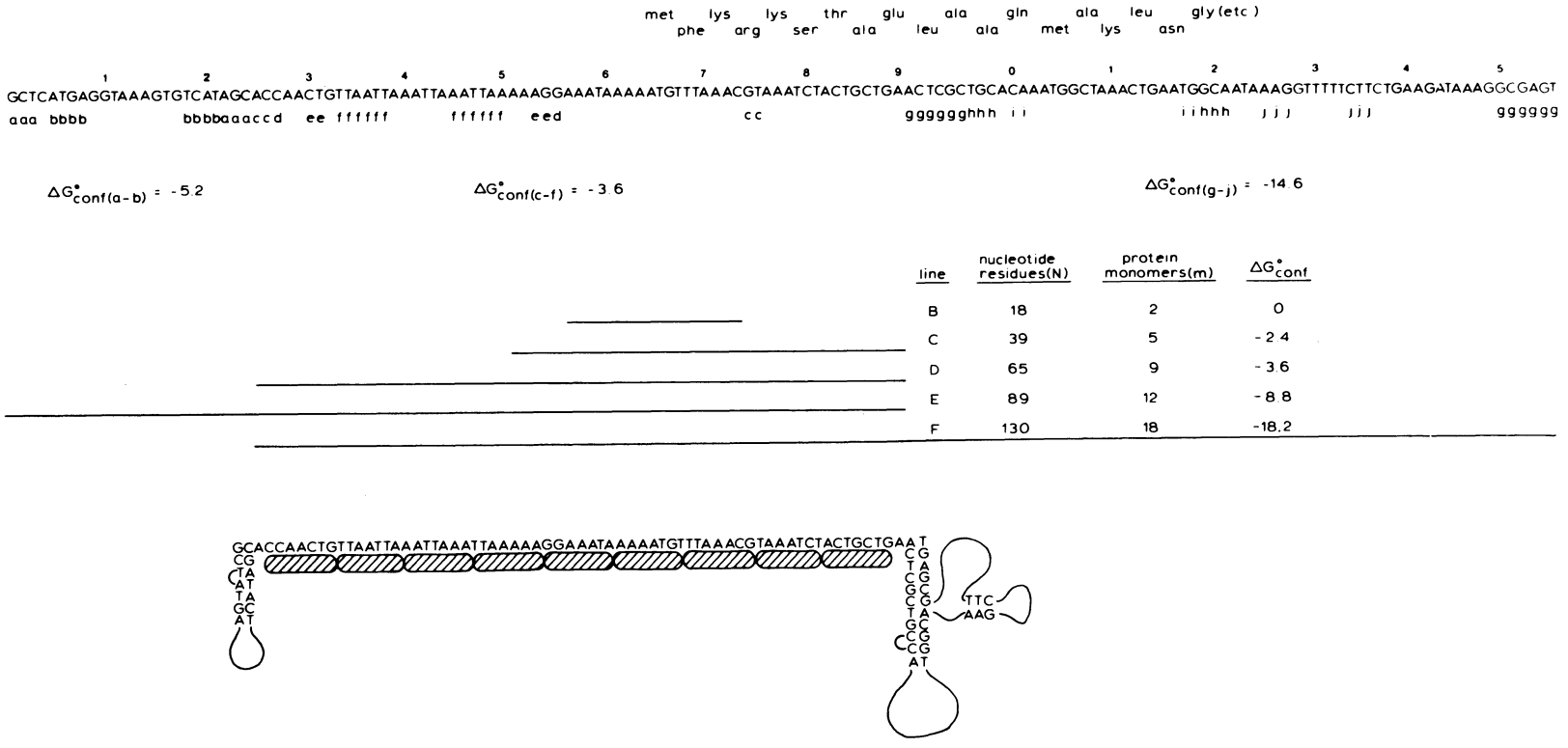

Fig. 4. Sequence and conformational stability of the putative gene 32 mRNA operator site and vicinity. At the top we show the DNA sequence (noncoding strand only; the sequence as written corresponds to mRNA when T is replaced by U), with the beginning of the gene 32 protein sequence written above it. The lower case letters below the DNA sequence correspond to possible base-pairing interactions; i.e., the bases marked aaa can pair with the subsequent aaa sequence to form the stem of a hairpin structure, etc. The lines (labelled B through F) correspond to the segments tested as potential operator sites (see text). The structure at the bottom is the preferred operator sequence, drawn in a gene 32 protein saturated conformation showing the proposed flanking hairpin termini. [Figure from Ref. 6].

even totally unstructured mRNA sequences (of average T4 composition) will not bind gene 32 protein under physiological conditions and protein concentrations if they are less than m = 4 (~28 nucleotide residues) in length. Furthermore, also due to this effect, even longer regions containing elements of weak secondary structure will remain uncomplexed. We expect that under physiological conditions the average mRNA molecule will be highly structured; thus regions sufficiently unstructured to bind gene 32 protein under intra-cellular conditions may be relatively rare.

The Gene 32 mRNA Translational Operator Site. The above calculations suggest that, in principle, the simplest way to define the gene 32 protein translational operator site, and to insure that it saturates at lower free gene 32 protein concentrations than do "control" sequences on other T4 mRNAs, is to have the gene 32 mRNA operator consist of a uniquely unstructured segment, as originally proposed by Russel et al. (Ref. 9). The combination of the availability of Kω values for all the relevant nucleic acid lattices, the sequencing of gene 32 mRNA by Krisch et al. (Ref. 7,18), and the availability of a large T4 DNA sequence library (Ref. 6), have now made it possible to test this suggestion quantitatively (Ref. 6).

The sequence surrounding the initiation codon of the gene 32 message is shown in Fig. 4. In most mRNA sequences this region contains the ribosomal binding site at which translation is initiated (Ref. 6), and thus comprises the most logical candidate for the gene 32 mRNA translational operator site. This view is based on the simplest translational repression model, in which gene 32 protein (as repressor) competes with the ribosome for this operator-initiator site.

The sequence of gene 32 mRNA in the vicinity of the initiation codon is remarkable, even for a phage containing 66% adenine plus thymine residues. As Fig. 4 shows, the ribosome binding site region contains a stretch of 40 nucleotides (residues 33 to 72, inclusive) in which the only nucleotides other than A or U are the three nearly essential G residues of the Shine-Dalgarno sequence and the initiation codon (see Ref. 6). Values of ΔG°_{conf} were computed for a variety of arbitrary segments within the gene 32 initiation sequence, in order to determine whether an unstructured domain of sufficient length to serve as an operator site could exist in this region within the quantitative constraints outlined above (for details see Ref. 6). Some of the results are shown in Fig. 4. In essence, it was found (Ref. 6) that the longest (unstructured and partially structured) potential operator sequence that can be saturated under intracellular conditions and at the regulated gene 32 protein concentration is represented by line D in Fig. 4. This sequence is shown in the bound conformation (complete with stable flanking hairpins) at the bottom of the figure; it binds nine gene 32 protein monomers.

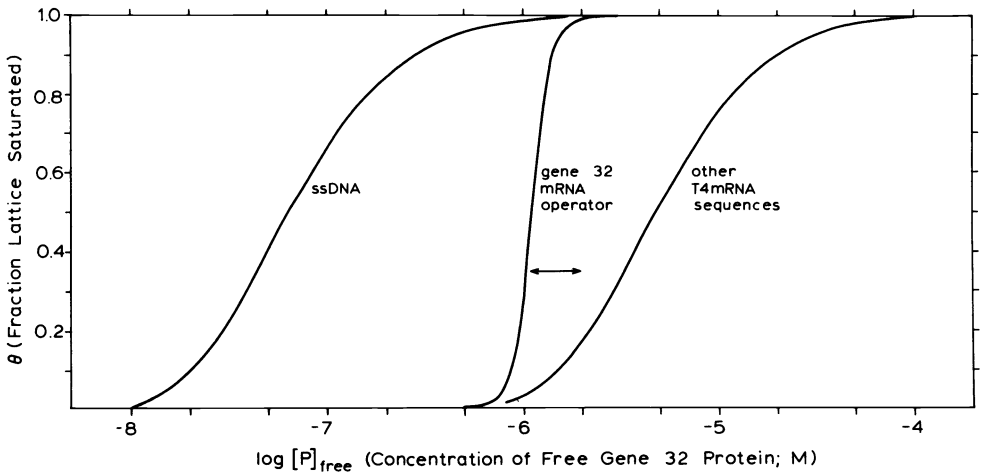

Fig. 5. Binding curves summarizing the gene 32 protein autoregulatory system. The "ssDNA" curve is calculated using the real T4 DNA sequences with an N = 50 residue lattice length replication window and the infinite lattice calculation mode. The "gene 32 mRNA operator" curve is calculated for the putative operator structure (line D) shown at the bottom of Fig. 4. The "other mRNA" curve is calculated using real T4 sequences with an N = 50 residue lattice length and the finite lattice approach. [Figure from Ref. 6.]

Is the Gene 32 mRNA Operator Sequence Unique? It was also, of course, necessary to determine whether the proposed gene 32 mRNA operator sequence defined in Fig. 4 is unique. To this end calculations were carried out using the entire catalog (Ref. 6) of T4 nucleic acid sequences (Ref. 6). The results showed that the proposed gene 32 mRNA operator has much less secondary structure than virtually any other sequences within the T4 sequence catalog (~5% of the total T4 genome). Comparison with more than 10 other T4 ribosome binding sites showed none to be as unstructured as the proposed gene 32 mRNA operator (Ref. 6).

The T4 Gene 32 mRNA Autogenous Regulatory System. The conclusions outlined above are summarized in Fig. 5, for the actual T4 system. Fig. 5 shows that, as required, the actual single-stranded DNA sequences of the T4 DNA replication complex (and presumably also of the T4 DNA recombination and repair systems) are saturated with gene 32 protein at concentrations well below the autoregulated value. The proposed gene 32 mRNA translational operator site then saturates quite sharply (cooperatively) at free protein concentrations just below the autoregulated level. As required, other T4 mRNA initiation (ribosomal binding) sequences are not appreciably complexed at the maintained intracellular free gene 32 protein concentration.

MODEL APPROACHES TO THE AUTOREGULATION OF SYNTHESIS OF PROTEINS INVOLVED IN THE ASSEMBLY OF THE E. COLI RIBOSOME

The autoregulation model developed for T4 gene 32 protein, as outlined above, is attractive and simple, and we have considered whether it might be modified to apply to the autoregulation of the synthesis of proteins involved in more complicated cellular systems as well. The self-assembly of the ribosome (Ref. 20) offers an intriguing case in point.

In outline, the E. coli ribosome consists of two subunits, containing a total of ~50 proteins (largely incorporated as one copy per ribosome), and 3 species of ribosomal RNA. Present views suggest that assembly involves the initial condensation of specific central proteins onto defined rRNA loci (defined, at least in part, by specific elements of rRNA secondary structure). These "primary assembly" proteins, together with certain portions of the surrounding rRNA, then create sites for the binding of "secondary" and "tertiary" assembly proteins, until the active ribosome is fully formed. Under exponential growth conditions the cell can contain as many as 10^5 ribosomes, at least 85% of which are fully functional, i.e. in polysomes (Ref. 22). This fact, in itself, means that the assembly process must be highly cooperative (see below). The problem for the cell is therefore to autoregulate the synthesis of the ~50 proteins of the ribosome so that ~10^5 copies of each are available at the proper concentrations and times in the cell cycle, to arrange that none are appreciably overproduced, and to assemble complete ribosomes in a cooperative manner.

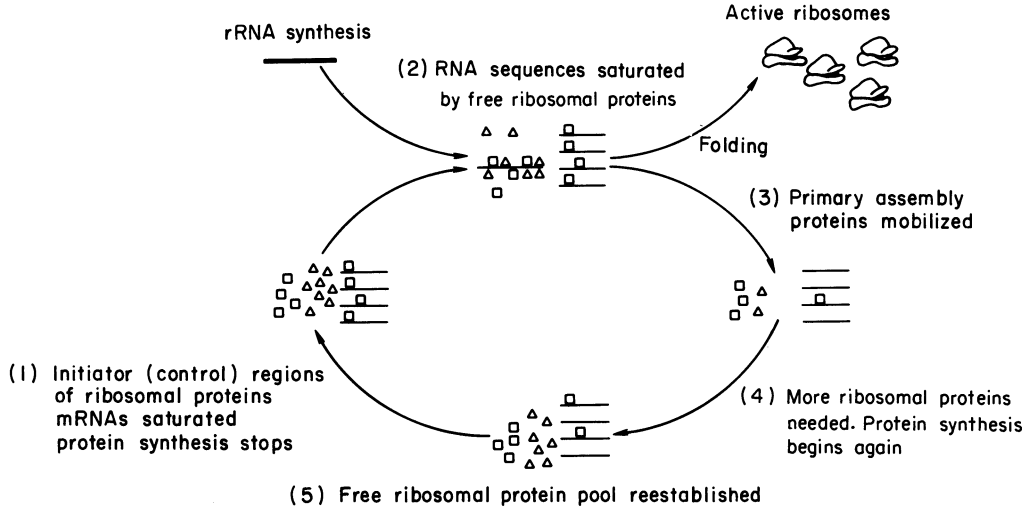

Fig. 6. Ribosomal Protein Autoregulatory Cycle: Schematic representation of the sequence of events as elucidated by Nomura and coworkers (Ref. 5). The various binding equilibria involved are indicated. The four lines drawn as a unit represent the ribosomal protein mRNAs. The heavy line at the top of the figure represents the ribosomal RNA. The squares denote the "primary assembly proteins". The triangles represent other ribosomal proteins.

In particular, Nomura and coworkers have shown that the genes for the ribosomal proteins are clustered into seven operons (regions of DNA under the control of one operator and transcribed as one mRNA), each coding for 2 to 11 ribosomal or related proteins (for a recent review see Ref. 5). Furthermore, one protein within each operon autogenously regulates synthesis from that operon. The regulated protein is generally, but not always, a "primary assembly" protein. In analogy to the gene 32 protein autoregulation system, these workers have put forward an elegant model in which the regulatory protein of each operon binds to its ribosomal target site (at least in part, a specific rRNA "hairpin" structure [in the cases that have been studied so far]). This regulatory protein also binds to an mRNA control site carrying a very similar hairpin structure (Ref. 23). Binding to the latter site coordinately turns off the synthesis of that operon. This model is presented schematically in Fig. 6.

This approach is clearly attractive in outline, but when considered in quantitative detail and in parallel with the gene 32 protein scheme, some problems emerge. However, these problems can be overcome; and in the final portion of this article, we describe some aspects of a quantitative model of how the autoregulatory aspects of such a "heteroprotein" self-assembly system might be viewed. Clearly the notions presented are general, and could apply to a variety of self-assembly systems. Elsewhere (Ref. 24) we have presented a more detailed treatment of the ribosome system *per se*.

Binding Models. The gene 32 protein system (see Fig. 4 and 5) works for three reasons: (i) only one kind of protein is involved; (ii) protein binding to the regulatory target (as well as to the functional target---the single-stranded DNA of the replication complex), is cooperative, and thus repression is an abrupt function of free gene 32 protein concentration; and (iii) the system takes advantage of the thermodynamics of finite lattice binding and thus creates binding specificity by utilizing as the autoregulatory target the longest "unstructured" mRNA translational initiation region present within the entire T4 mRNA constellation (see Fig. 4).

None of these conditions apply to the ribosomal binding problem. Here: (i) binding of a number of *different* proteins is involved; (ii) binding to both the functional (rRNA) and the regulatory (mRNA) targets is "single-copy", and thus, by definition, cannot be cooperative in the concentration of *that individual protein*; and (iii) the functional binding, and presumably the autoregulatory binding, are to structured rather than to unstructured RNA sequences.

The quantitative problems associated with an autoregulatory system involving noncooperative functional binding of a given single ribosomal protein (arbitrarily designated P1) to an rRNA target (e.g., a specific hairpin), followed (at the same or a higher concentration of free P1) by noncooperative binding to an autoregulatory mRNA target (e.g., a hairpin of comparable structure), are shown in curve (a) of Fig. 7.

First, since binding is noncooperative, saturation (from $f_{complex} \simeq 0.05$ to $\simeq 0.95$, see below) of the functional (and the autoregulatory) target(s) will proceed over a 1000-fold change in free P1 concentration. This requires either that binding be very tight (i.e., that both $K_{1,rRNA}$ *and* $K_{1,\ mRNA}$ be large) and thus that most of the synthesis of ribosomal proteins proceeds primarily under largely repressed conditions ($f_{mRNA} > 0.9$), or that the system operate at very large excesses (up to 1000-fold) of ribosomal proteins beyond the amounts required to assemble the ribosomes themselves. Neither of these possibilities applies *in vivo*: Maaloe (Ref. 25) and others have shown that ribosomal proteins are synthesized at rates that, at least, come close to the constitutive limit, and thus it seems unlikely that this synthesis proceeds under "throttled down" conditions; and free ribosomal proteins are not present in appreciable excess in the functional *E. coli* cell.

How, then, might assembly be driven to completion at low free concentrations of ribosomal proteins, and "shut-off" be completed without production of large quantities of excess ribosomal proteins? Several levels of model solutions are illustrated in Fig. 7.

Two Proteins, One rRNA Site. Since, in effect, only one (regulatory) protein binds to the translational shut-off site on each coordinately regulated messenger RNA, not much (in a thermodynamic sense) can be done at the level of the regulatory binding, and the solution to the problem must be found in the multi-protein interactions involved in binding to the functional target. That is, the binding of subsequent proteins to the primary assembly complexes on the ribosome must effectively "sharpen up" the rRNA titration curve, and/or shift it toward lower free (P1) protein concentrations. These notions can be clearly modelled with a two protein-one rRNA site system.

In this simple system, protein 1 (P1) binds directly to the rRNA site, and protein 2 (P2) then binds to the resulting complex. (P2 has little affinity for either the rRNA site, or for P1, separately.) The titration labelled curve (a) in Fig. 7 plots the binding of [P1] alone, either to the functional (rRNA) or to the regulatory (mRNA) target. Thus, if a single copy of the control protein (P1) binds to the regulatory target, curve (a) is the

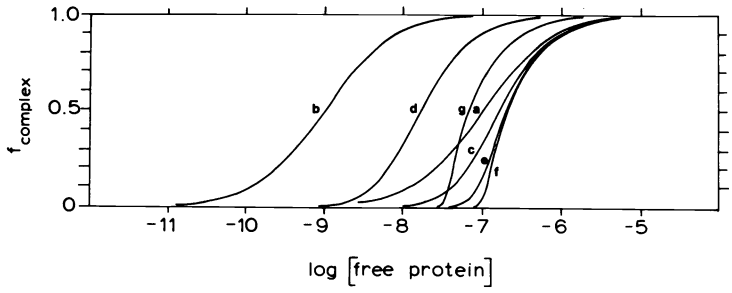

Fig. 7. Binding curves summarizing multiple protein binding to an rRNA locus. The fraction of RNAs covered by the complete complex, $f_{complex}$, as a function of the free concentration of the first binding protein is shown. The rRNA concentration is assumed to be constant at $1x10^{-7}$ M. a - Binding of one protein to its target site on mRNA or rRNA with a binding constant of $1x10^{7}$ M^{-1}. b - Binding of first protein to form a two protein complex in the presence of $1x10^{-7}$ M of a more tightly binding second protein. $K_1 = 1x10^{7}$ M^{-1}, $K_2 = 1x10^{9}$ M^{-1}. c - Cooperative binding of two proteins. Both binding constants are $1x10^{7}$ M^{-1}. The concentrations of the two proteins are equal. d - Cooperative binding of two dissimilar proteins. The second protein binds more strongly as in b, but now the concentrations of the two proteins are always equal. e - Cooperative binding of five proteins. Equal concentrations of all proteins. All binding constants are 10^{7} M. f - Cooperative binding of twenty proteins assemblying onto one site, equal concentrations of all proteins. g - Tighter cooperative binding of twenty proteins assemblying onto one site, equal concentrations of all proteins. The binding constant of all proteins except protein 1 is 3 $1x10^{7}$ M^{-1} (see text for further details).

"shut-off" curve. We use a binding constant (K_1) of $10^{7}M^{-1}$ for this interaction, both because this is close to the values actually measured for the binding of some single ribosomal proteins to their rRNA targets (Ref. 26), and because this represents a binding affinity at which metastable complexes can "anneal" to their final forms at biologically appropriate rates. Thus this titration effectively goes to completion at a free P1 concentration of $\sim 10^{-6}$M, corresponding to $\sim 10^{3}$ molecules of free P1 per cell. (This level of regulatory synthesis requires only a 1% overproduction of P1, since up to $\sim 10^{5}$ molecules of P1 have been incorporated into ribosomes.) The equilibrium constants (K_1 and K_2) for this two-step process are:

$$P1 + rRNA = P1 \cdot rRNA \qquad K_1 = \frac{[P1 \cdot rRNA]}{[P1]\ [rRNA]} \tag{4}$$

$$P2 + P1 \cdot rRNA = P2 \cdot P1 \cdot rRNA \qquad K_2 = \frac{[P1 \cdot P2 \cdot rRNA]}{[P2]\ [P1 \cdot rRNA]} \tag{5}$$

where P1 = protein 1, P2 = protein 2, P1•RNA = the complex of protein 1 and the rRNA and P1•P2•RNA = the complex of protein 1, protein 2, and the rRNA site.

For the assembly of multiprotein complexes, θ (eq. 1) can be replaced by two particularly useful measures, the fraction of the total RNA that has the first protein bound to it (f_{RNA}), and the fraction of the total RNA that has been assembled into a complete complex ($f_{complex}$). For the assembly of partial or complete ribosomes, f_{RNA} for the binding of one protein can also be used to describe the level of complexation of the regulatory target site (curve a of Fig. 7; see above).

For the two protein-one rRNA site system, f_{RNA} and $f_{complex}$ have the following simple forms:

$$f_{RNA} = \frac{K_1\ [P1] + K_1K_2\ [P1]\ [P2]}{1 + K_1\ [P1] + K_1K_2\ [P1]\ [P2]} \tag{6}$$

and

$$f_{complex} = \frac{K_1K_2\ [P1]\ [P2]}{1 + K_1\ [P1] + K_1K_2\ [P1]\ [P2]} \tag{7}$$

For ribosomal proteins and RNA, P1 is used to designate an autoregulatory protein that shuts off its own synthesis. The behavior of f_{RNA} and $f_{complex}$ have therefore been investigated under two conditions to see whether the presence of secondary binding proteins facilitates the assembly of ribosomes before P1 synthesis is suppressed. First, the effect of a constant concentration of P2 on the assembly of P1 onto rRNA and on complete complex formation is studied (i.e.; P1 is synthesized in the presence of a constant concentration of P2). Then we model the effect of coupling the concentrations of P1 and P2 (i.e.; both proteins are produced coordinately from the same autoregulated mRNA), so that as the concentration of one protein increases that of the other does also.

In curve (b) of Fig. 7, we show that increasing K_2 one hundred-fold above K_1 results in an effective one hundred-fold increase in the net binding constant of P1 for rRNA; i.e., the titration curve for this binding is shifted to the left by 2 log units, with no sharpening of the rRNA titration curve. This occurs because (through K_2) the concentration of completed complex dominates the numerator of eq. (6), and thus $f_{RNA} \simeq f_{complex}$.

Further examination of the relationships of f_{RNA} and $f_{complex}$ to [P1] shows that the concentration of free P1 required to half-titrate either the RNA or the complete complex is mediated not by K_2 alone, but by the product $K_2[P2]$. The contribution of the concentration of P2 in changing the apparent binding constant of P1 will be seen to be crucial for the coordinated control of ribosomal synthesis. We note that the participation of constant levels of P2 in the binding of P1 to RNA therefore does not sharpen the binding curve; i.e., formation of the complex is not cooperative.

When the levels of free P1 and free P2 in the cell are coupled (and held equal), the equations for f_{RNA} and $f_{complex}$ can be written:

$$f_{RNA} = \frac{K_1 [P1] + K_1K_2 [P1]^2}{1 + K_1 [P1] + K_1K_2 [P1]^2} \tag{8}$$

$$f_{complex} = \frac{K_1K_2 [P1]^2}{1 + K_1 [P1] + K_1K_2 [P1]^2} \tag{9}$$

Now a new behavior is seen in the relationship between $f_{complex}$ and [P1]; see curves (c) and (d) of Fig. 7. Here the titration occurs over a narrowed range of [P1], since it has become cooperative in the product of the concentrations of [P1] and [P2] (equal to $[P1]^2$ in eqs. 8 and 9, since [P1] = [P2]). In other words, because of the coordinate regulation of the free concentrations of P1 and P2, the assembly of the P1•P2•rRNA complex has become cooperative. Because two different proteins are involved we call this heteroprotein cooperativity. If $K_2 = K_1$, the assembly titration is sharpened (curve c). If $K_2 > K_1$, the assembly titration is shifted and sharpened (curve d). In either case assembly can now go to completion before shut off of synthesis by the binding of P1 to the mRNA target site.

Coupling of Initial Assembly Complexes. Of course, many more complex models can be envisioned, and probably apply in actual ribosome assembly (see Ref. 24 for details). One example is labelled curve (e) in Fig. 7. Here the sequential assembly of two "1 rRNA + 2 protein" complexes, as outlined above, are coupled by a "bridging" fifth protein that, effectively, ties together the assembly of the complexes originating on two independent loci of the same rRNA. All proteins are coordinately regulated (i.e., [P1] = [P2] = ...[P5]), and $K_1 = K_2 = \ldots K_5$. Thus titration is effectively cooperative in the fifth-power of [P1] (see below).

Assembly of an Entire Ribosomal Subunit. Eqs. (3) and (4) have been extended to the assembly of complexes containing any number of proteins in eq. (10) (Ref. 24). Even though the general form of this n protein equation is complicated, the cooperative nature of the assembly of the multiprotein complex is made clear by assuming that all of the proteins are coordinately regulated to identical free concentrations and that the proteins have identical binding constants to their appropriate portions of the assembling complex. $f_{complex}$ now assumes the following simple form:

$$f_{complex} = \frac{(KP)^n}{1 + \sum_{i=1}^{n} (KP)^i} \tag{10}$$

where P is concentration of one of the free proteins, K is its binding constant, and n is the number of proteins in this complex. Eq. 10 demonstrates the essential feature of the cooperative production of a multiprotein complex. The nth degree polynomials in the numerator and the denominator narrow the free protein concentration range over which formation of complete complex occurs. Thus the formation of complete complex is cooperative in the free protein concentration; i.e., heteroprotein cooperativity exists. Curves (f) and (g) show such behavior for the coordinated assembly of the small ribosomal subunit (with n = 20), for $K_1 = K_2 = K_3 \ldots K_{20}$ (curve f) and for $K_2 = 3\ K_1$ and $K_2 = K_3 = K_4 \ldots K_{20}$ (curve g).

Heteroprotein Assembly Processes. The above results, using simple model systems, show that the specific features of the gene 32 protein autoregulatory system can be generalized to provide quantitatively workable translational repression schemes for much more complex systems. Elsewhere (Ref. 24), we have shown how these approaches can explain quantitatively several features of the physiology and genetics of E. coli ribosomes, including the paucity of incomplete ribosomes and the low level of overproduction of ribosomal proteins, the ordering of the various ribosomal proteins of the small and large ribosomal subunits within the ribosomal protein operons (this ordering effectively coordinates the synthesis of all the ribosomal proteins, and allows assembly of the large subunit to keep pace with that of the small subunit), etc. It may well be that these notions can also be applied to the regulation of the synthesis of the components of a variety of other cellular organelles. Time, and further experimentation, will tell.

Acknowledgments. We are grateful to our colleagues Stephen Kowalczykowski, Nils Lonberg, John Newport and Leland Paul for their contributions to the physical chemical elucidation of the gene 32 protein autoregulation system as described here, as well as to Gary Stormo for computer modelling of the T4 sequence conformations (Fig. 4 and 5), and to Larry Gold for many stimulating discussions and written interactions on autoregulatory systems. Their contributions are clearly delineated by the authorships of the various papers reviewed here. The research work described here was supported in part by USPHS Research Grants GM-15792 and GM-29159 (to PHvH), and by USPHS Postdoctoral Fellowship GM-08252 (to FRF).

REFERENCES

1. G.R. Stark and G.M. Wahl, Ann. Rev. Biochem. 53, 447-92 (1984).
2. P.H. von Hippel, D.G. Bear, W.M. Morgan, and J.A. McSwiggen, Ann. Rev. Biochem. 53, 389-446 (1984).
3. D.H. Doherty, P. Gauss, and L. Gold, in Multifunctional Proteins: Catal./Struct. Regul. (J.F. Kane, ed) CRC Press, Boca Raton, FL, 45-72 (1983).
4. K.M. Campbell, G.D. Stormo, and L. Gold, in Gene Function in Prokaryotes (J. Beckwith, J. Davies, and J.A. Gallant, eds) Cold Spring Harbor Laboratory, NY, 185-210 (1983).
5. M. Nomura, R. Gourse, and G. Baugham, Ann. Rev. Biochem. 53, 75-118 (1984).
6. P.H. von Hippel, S.C. Kowalczykowski, N. Lonberg, J.N. Newport, L.S. Paul, G.D. Stormo, and L. Gold, J. Mol. Biol. 162, 795-818 (1982).
7. H.M. Krisch, R.M. Duvoisin, B. Allet, and R.H. Epstein, ICN-UCLA Symp. on Mol. and Cell Biol. 29, 517-526 (1980).
8. L. Gold, P.A. O'Farrell, and M.L. Russel, J. Biol. Chem. 251, 7251-7262 (1976).
9. M.L. Russel, L. Gold, H. Morrissett, and P.Z. O'Farrell, J. Biol. Chem. 251, 7263-7270 (1976).
10. G. Lemaire, L. Gold, and M. Yarus, J. Mol. Biol. 126, 73-90 (1978).
11. J. Karam, L. Gold, B.S. Singer, and M. Dawson, Proc. Natl. Acad. Sci. USA 78, 4669-4673 (1981).
12. D.E. Jensen, R.C. Kelly, and P.H. von Hippel, J. Biol. Chem. 251, 7215-7228 (1976).
13. J.W. Newport, N. Lonberg, S.C. Kowalczykowski, L.S. Paul, and P.H. von Hippel, J. Mol. Biol. 145, 105-121 (1981).
14. R.C. Kelly, D.E. Jensen, and P.H. von Hippel, J. Biol. Chem. 251, 7240-7250 (1976).
15. S. Kowalczykowski, N. Lonberg, J.W. Newport, and P.H. von Hippel, J. Mol. Biol. 145, 75-104 (1981).
16. J.D. McGhee and P.H. von Hippel, J. Mol. Biol. 86, 469-489 (1974).
17. Y. Kao-Huang, A. Revzin, A.P. Butler, P. O'Connor, D. Noble, and P.H. von Hippel, Proc. Natl. Acad. Sci. USA 74, 4228-4232 (1977).
18. H.M. Krisch and B. Allet, Proc. Natl. Acad. Sci. USA 79, 4937-4941 (1982).
19. I. Tinoco, P.N. Borer, B. Dengler, and J. Gralla, Nature New Biol. 246, 40-41 (1973).
20. M. Nomura and W.A. Held, in Ribosomes, Cold Spring Harbor Laboratory, New York, 193-224 (1974).
21. K.H. Neirhaus, in Ribosomes: Structure, Function, and Genetics (G. Chambliss, G.R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura, eds.) University Park Press, Baltimore, 267-294 (1980).
22. K. Gausing, Mol. Gen. Genet. 129, 61-75 (1974).
23. J. Yates and M. Nomura, in Multifunctional Proteins: Catal./Struct. Regul. (J.F. Kane, ed) CRC Press, Boca Raton, FL, 31-43 (1983).
24. F.R. Fairfield and P.H. von Hippel, Proc. Natl. Acad. Sci. USA, in press.
25. O. Maaløe, in Biological Regulation and Development (R.F. Goldberger, ed) Plenum Press, New York, 487-542 (1979).
26. J. Schwarzbauer and G.R. Craven, Nuc. Acids Res. 9, 2223-2237 (1981).
27. P.H. von Hippel and F.R. Fairfield, in Mobility Recognit. Cell Biol. (H. Sund and C. Veeger, eds) W. de Gruyter, Berlin, 213-236 (1983).