Interactions of Bacteriophage T4-coded Gene *32* Protein with Nucleic Acids

II. Specificity of Binding to DNA and RNA

John W. Newport†, Nils Lonberg‡, Stephen C. Kowalczykowski
and Peter H. von Hippel

Institute of Molecular Biology
and
Department of Chemistry
University of Oregon
Eugene, Ore. 97403
U.S.A.

In this paper we examine the specificity of the co-operative binding (in the polynucleotide mode) of bacteriophage T4-coded gene *32* protein to synthetic and natural single-stranded nucleic acids differing in base composition and sugar type. It is shown by competition experiments in a tight-binding (low salt) environment that there is a high degree of binding specificity under these (protein-limiting) conditions, with one type of nucleic acid lattice binding gene *32* protein to saturation before any binding to the competing lattice takes place; it is also shown that the same differential specificities apply at high salt concentrations. Procedures developed in the preceding paper (Kowalczykowski *et al.*, 1980) are used to measure the net binding affinities ($K\omega$) of gene *32* protein to a variety of polynucleotides, as well as to determine individual values of K and ω for some systems. For all polynucleotides, virtually the entire specificity and salt dependence of binding of $K\omega$ appears to be in K. In ~ 0.2 M-NaCl, the net binding affinities ($K\omega$) range from $\sim 10^6$ to $\sim 10^{11}$ M^{-1}; in order of increasing affinities we find: poly(rC) $<$ poly(rU) $<$ poly(rA) $<$ poly(dA) $<$ poly(dC) $<$ poly(dU) $<$ poly(rI) $<$ poly(dI) $<$ poly(dT). In general, $K\omega$ for a particular homopolyribonucleotide at constant salt concentration is 10^1 to 10^4 *smaller* than $K\omega$ for the corresponding homopolydeoxyribopolynucleotide. Values of $K\omega$ for randomly copolymerized polynucleotides and for natural DNA fall at the compositionally weighted average of the values for the individual homopolynucleotides (except for poly(dT), which appears to bind somewhat tighter), indicating that the net affinity represents the *sum* of the binding free energy contributions of the individual nucleotides. It is shown that these results, on a competition basis under physiological salt conditions, can account quantitatively for the autogenous regulation of the synthesis of gene *32* protein at the translational level (Russel *et al.*, 1976; Lemaire *et al.*, 1978). In

† Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 994143, U.S.A.

‡ Present address: Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138, U.S.A.

addition, these results suggest possible mechanisms by which gene *32* messenger RNA might be specifically recognized (by gene *32* protein) and functionally discriminated from the other mRNAs of phage T4.

1. Introduction

In the Introduction to the preceding paper (Kowalczykowski *et al.*, 1980) we discussed the biological properties of gene *32* protein as determined in studies *in vivo* and *in vitro*, and pointed out that the specificity requirements of the autoregulatory control of the synthesis of this protein (Krisch *et al.*, 1974; Russel *et al.*, 1976; Lemaire *et al.*, 1978) appeared to be incompatible with the results of earlier physico-chemical studies from this laboratory (Kelly *et al.*, 1976). The essence of the problem seemed to be to develop a molecular explanation for the hierarchy of biologically observed gene *32* protein binding specificities (with relative affinities decreasing as follows: single-stranded DNA sequences > gene *32* protein messenger RNA > other bacteriophage T4 mRNAs > double-stranded DNA, etc.) in light of the apparent non-specificity of binding demonstrated with the oligonucleotides.

In principle the problem is soluble at this level; von Hippel *et al.* (1977; see also below) had pointed out that even very small differences (within the limits of error of the absolute binding measurements reported by Kelly *et al.* (1976)) in the binding affinity of individual gene 32 protein molecules for the various short oligonucleotide lattices could be appreciably amplified (to approximately the power of the average cluster size) if the protein binds co-operatively.

To attempt a quantitative approach to this problem, it was necessary first to redetermine and to extend, at the highest attainable accuracy, the binding measurements of the various oligonucleotides. These results, summarized in the preceding paper (Kowalczykowski *et al.*, 1980), confirmed the earlier conclusions (Kelly *et al.*, 1976) that various A and T-containing oligonucleotides bind to gene *32* protein with comparable affinities, and showed that the affinities of gene *32* protein for G and C-containing species are also about the same. Furthermore, ribose-containing oligonucleotides seemed to bind to gene *32* protein only marginally more weakly than their deoxyribose-containing homologues. Thus if the physiologically relevant binding hierarchy listed above was indeed to find its explanation *via* direct cluster-based amplification of the virtually indistinguishable binding affinities of the oligonucleotides for gene 32 protein, discrimination would have to be subtle indeed.

However, in the course of studying the co-operative binding (and particularly the salt dependence of this binding) of gene *32* protein to polynucleotides, we found that several aspects of this binding differed qualitatively from that of the short oligonucleotides studied previously, suggesting that two different binding conformations are involved. This led us to make a careful quantitative examination of the molecular details of the polynucleotide binding interaction, and in the preceding paper we developed a detailed model for the interactions of gene *32* protein with nucleic acids in both binding conformations. These findings make it possible for us to ask again how physiological binding specificity develops; this time in the context of the differences in the binding affinity of the protein interacting with nucleic acid lattices in the *polynucleotide* binding mode.

In this paper we report an extensive series of measurements on the affinity of gene *32* protein for polynucleotides of varying base and sugar composition. The results show that differential (specific) binding affinities do exist at this level and, together with binding co-operativity, do indeed lead to a simple physico-chemical model that can account for the physiological binding and autoregulatory properties of the system.

2. Materials and Methods

Most materials, buffers and methods used in this study are described in the preceding paper. Small differences in procedure are described in Results. The sources and extinction coefficients of all the polynucleotides used in these studies are listed in the accompanying papers (Kowalczykowski *et al.*, 1980; Lonberg *et al.*, 1980).

3. Results

(a) *Specificity of gene* 32 *protein binding to polynucleotides*

In the preceding paper (Kowalczykowski *et al.*, 1980) we showed that the binding affinity of gene *32* protein for short ($l=2$ to 8) oligonucleotides is essentially independent of the nucleotide composition of the nucleic acid lattice. We also showed that there are marked differences between the binding behavior (as reflected in K) of gene *32* protein in the oligo- and in the polynucleotide binding modes. In this paper we ask whether binding in the co-operative polynucleotide mode shows specificity on the basis of either base composition or sugar type.

(b) *Binding competition experiments demonstrate specificity at low salt concentrations*

Under low salt conditions (~ 10 mM-Na$^+$), gene *32* protein has been shown to bind tightly and co-operatively to poly(dA) and poly(rA). A large hyperchromicity is seen at ~ 260 nm when protein binds to single-stranded base-stacked polynucleotides, corresponding to base unstacking and backbone deformation. By following this hyperchromicity as a function of added gene *32* protein (at a fixed concentration of polynucleotide), Jensen *et al.* (1976) established a site size (n) for the co-operative binding of gene *32* protein (to poly(dA)) of 7 (± 1) nucleotide residues; and Kowalczykowski *et al.* (1980) showed that the net binding affinity (to poly(rA)) is very salt concentration-dependent. Here we extend this approach to measure the *relative* binding affinities of co-operatively bound gene *32* protein for polynucleotides of varying base and sugar composition.

Solutions of single-stranded polynucleotides in 10 mM-NaCl, 1·0 mM-Na$_2$HPO$_4$ (pH 7·7) (buffer B; Kowalczykowski *et al.*, 1980) have been titrated with gene *32* protein, and the change in absorbance monitored at 260 nm. A typical low salt titration curve (for poly(rA)) is shown in Figure 1; note the expected sharp break at a stoichiometry of ~ 7 nucleotide residues per gene *32* protein monomer added, showing that binding saturation has been achieved. Similar titrations have been carried out with a variety of other polynucleotides; all reached binding saturation at
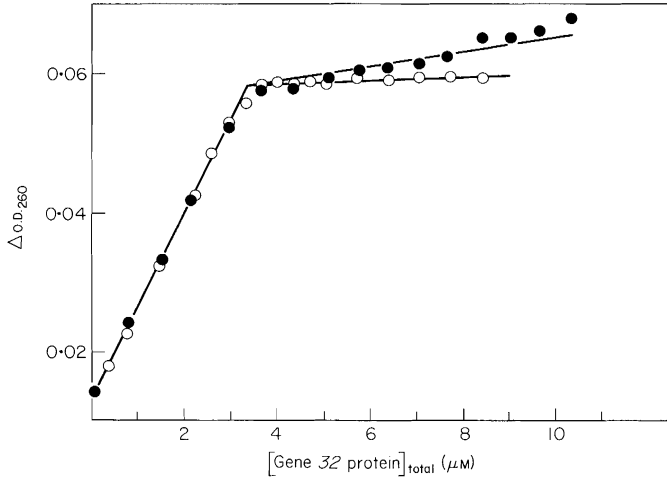
Fɪɢ. 1. Low salt competition titration of poly(rA) and poly(rC) in buffer B plus 10 mM-NaCl, plus the indicated concentrations of homopolynucleotides at 25°C. The change in absorption at 260 nm due to the binding of gene *32* protein was monitored as a function of protein added for: poly(rA) alone $(2.46 \times 10^{-5}$ M) (○); and for poly(rA) $(2.46 \times 10^{-5}$ M) plus poly(rC) $(2.5 \times 10^{-4}$ M) (●).

Tᴀʙʟᴇ 1

Fractional change in polynucleotide absorbance (at 260 nm) on saturation with gene
32 protein at low salt

Polynucleotide	Δ O.D.$_{260}$§
Poly(dA)	$+0.12 \pm 0.01$
Poly(rA)	$+0.17 \pm 0.02$
Poly(dC)	$+0.05 \pm 0.02$
Poly(rC)	$+0.05 \pm 0.02$
Poly[r(A,C)]†	$+0.07 \pm 0.01$
Poly(dT)	-0.05 ± 0.01
Poly(rU)	-0.03 ± 0.01
Poly(rA)-pretreated with formaldehyde‡	-0.03 ± 0.01

Measured in buffer B containing 10 mM-NaCl; optical densities have been corrected for protein absorbance and dilution effects.

† A to C ratio for this random copolymer is 0·83.

‡ Poly(rA) was incubated for several days in 5 M-HCHO at room temperature; just before use the free formaldehyde was removed by centrifugation of the solution through a small column as described by McGhee & von Hippel (1977). Titrations were completed within 40 min after this step. Under these circumstances at least 80% of the exocyclic amino groups of poly(rA) remain complexed as hydroxymethylol adducts (McGhee & von Hippel, 1977).

§ + =hyperchromic change; − =hypochromic change; Δ O.D.$_{260}$ =(O.D.$_{\text{final}}$ − O.D.$_{\text{initial}}$/O.D.$_{\text{initial}}$).

$n = 7$ (± 1) nucleotide residues under these conditions (see Table 1). However, both the *magnitude* and the *direction* of the change in apparent optical density of the polynucleotides on binding to gene *32* protein varies with base composition and sugar type. The values of Δ O.D.$_{260}$ determined for the various polynucleotides

examined are summarized in Table 1. We note that poly(dA) and poly(rA) show by far the greatest hyperchromic change, while a small *hypo*chromic change is seen with poly(dT), poly(rU) and formaldehyde-treated poly(rA).

We have exploited these differences in optical density change on binding of gene *32* protein to different polynucleotides in competition studies to examine the relative affinities of this protein for different polynucleotide lattices under co-operative binding conditions. A typical experiment is shown in Figure 1, where the titration of poly(rA) is monitored in the absence and in the presence of a tenfold molar excess of poly(rC). Both polynucleotides bind gene *32* protein stoichiometrically under these low salt conditions, yet Figure 1 clearly shows that in this competition experiment all the poly(rA) is titrated to completion before the titration of *any* of the poly(rC). Thus the affinity of gene *32* protein, binding in the co-operative polynucleotide mode, is appreciably greater for poly(rA) than for poly(rC); resulting in appreciable *apparent* binding specificity under these (protein-limiting) conditions. The physico-chemical basis of such apparent differential binding specificities will be considered further in the Discussion.

Such competition experiments have been carried out with other pairs of polynucleotides (avoiding pairs capable of inter-chain base-pairing), and the results have been used to establish a partial hierarchy of relative polynucleotide binding affinities for gene *32* protein. Because of the large differences in the gene *32* protein-induced hyperchromicity of poly(rA) and poly(dA), relative to the other poly-nucleotides tested, most pairwise experiments were conducted against one of these adenine-containing polynucleotides. We have shown that the binding affinity of gene *32* protein for poly(rA) at low salt concentrations is approximately the same as for formaldehyde-treated poly(rA) (carrying hydroxymethylol adducts on the N^6-amino group of the adenine moiety; see Table 1) and greater than the affinity of the protein for poly(rC) and poly[r(A,C)]. Additional experiments have established that the apparent affinity of gene *32* protein for poly(dA) exceeds that for poly(rC). Similar relative affinities for gene *32* protein binding to polynucleotides have been demonstrated by Bobst & Pan (1975), using spin-labelled poly(rA) as a probe in qualitative competition experiments of this type.

Obviously the above method for determining binding affinities, while providing a useful demonstration of specificity and a model of possibly physiologically relevant control systems based on binding competition for limiting protein molecules (see Discussion), is of only limited use as an analytical technique. Previously (e.g. see Fig. 3 of Kowalczykowski *et al.*, 1980) we had shown that binding of gene *32* protein to polynucleotides is appreciably weakened at high salt concentrations; an appreciable "lag" is seen in the co-operative binding isotherm, showing that the binding is *not* stoichiometric under those conditions. Thus the basic requirement for measuring binding constants, that concentrations of *both* bound and free ligand be measurable at equilibrium, are met under higher salt conditions, and values of K and ω can, in principle, be determined for these systems.

First, however, we must demonstrate that the differential binding specificities seen in the low salt competition experiments (Fig. 1) carry over to the higher salt concentrations. Figure 2 shows a high salt competition experiment in which poly(rA) is titrated in the presence and absence of a tenfold excess of the random
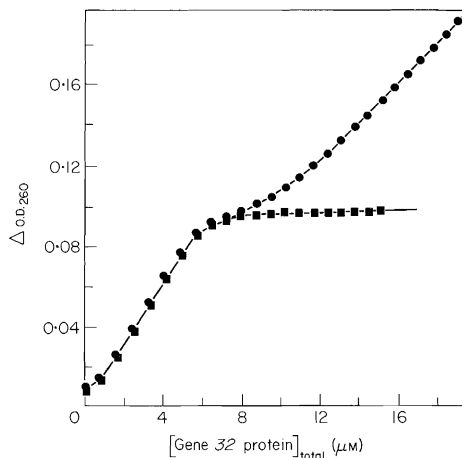
FIG. 2. u.v. absorbance titration with gene *32* protein of poly(rA) alone ($4 \cdot 13 \times 10^{-5}$ M) (■) or poly(rA) ($4 \cdot 13 \times 10^{-5}$ M) in the presence of a 10-fold excess of poly[r(A,C)] ($4 \cdot 2 \times 10^{-4}$ M) ($0 \cdot 83$ mol A/mol total nucleotide) (●). The titration was carried out in buffer B plus $0 \cdot 35$ M-NaCl.

copolymer, poly[r(A,C)]. Clearly, as before, poly(rA) is titrated to completion before *any* titration of poly[r(A,C)] occurs.

(c) *Absolute measurement of polynucleotide binding affinities for gene 32 protein*

(i) *Ultraviolet light absorbance measurements*

As described by Kowalczykowski *et al.* (1980), the absolute binding affinity ($K\omega$) of a polynucleotide for gene *32* protein can be measured under conditions of increased salt concentration where binding is *not* stoichiometric. Using this procedure, ultraviolet light absorbance titrations of various polynucleotides have been conducted, and values of $K\omega$ determined by measuring the *free* protein concentration (L_F) at the midpoint of the titration. The midpoints of the titration curves have been determined using the saturation values of hyper- or hypochromicity listed for each polynucleotide in Table 1, and a site size (n) of seven nucleotide residues per gene *32* protein monomer.

In order to determine reasonably accurate values of $K\omega$, an appreciable lag phase (e.g. see Fig. 3 of Kowalczykowski *et al.*, 1980) is required in the titration. In practice, this corresponds to a minimum value of L_F of $\sim 1 \times 10^{-6}$ M-gene *32* protein, and establishes the low salt concentration limit at which titrations to determine $K\omega$ can be carried out. The high salt concentration limit is set by the maximum concentration of free gene *32* protein that can, in practice, be present in a cuvette without excessive interference from light-scattering and aggregation; this limit is $\sim 2 \times 10^{-5}$ M-gene *32* protein.

We found that the salt concentrations over which titrations can be conducted on the above basis differ greatly from one polynucleotide to another. Thus values of $K\omega$ are conveniently measured between $0 \cdot 35$ and $0 \cdot 5$ M-NaCl for poly(rA), while measurable values of $K\omega$ for poly(dT) can be attained only at salt concentrations in excess of $1 \cdot 5$ M. Titrations were conducted by the ultraviolet light method at several
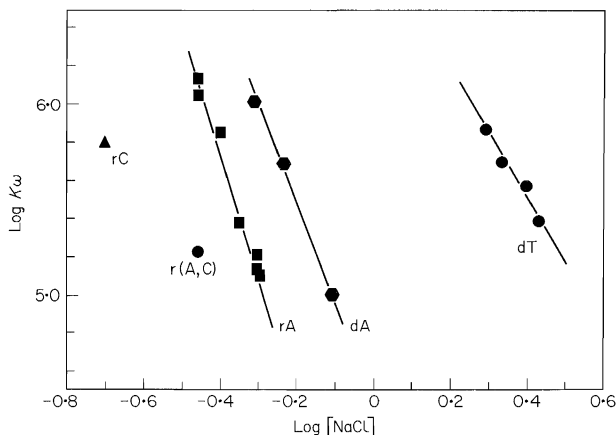
FIG. 3. Absolute values of $K\omega$ at different NaCl concentrations. A series of titrations under conditions as in Fig. 2, but at different NaCl concentrations, were carried out using poly(rC) (▲), poly[r(A,C)] (●), poly(rA) (■), poly(dA) (◆) and poly(dT) (●). Values of $K\omega$ were determined for each polynucleotide at each salt concentration as described in the text.

salt concentrations for several polynucleotides. The results are summarized as plots of log $K\omega$ *versus* log [NaCl] in Figure 3 and show, at least for poly(dA), poly(rA) and poly(dT), that log–log plots for these systems are quite linear, with rather similar slopes. Furthermore, these data confirm that the relative affinities of binding to the various polynucleotides fall in the order inferred from the competitive binding experiments (Figs 1 and 2). In addition (at least for adenine-containing poly-nucleotides), these results show that the affinity of gene *32* protein for poly-*deoxyribo*nucleotides exceeds that for the homologous poly*ribo*nucleotides.

Binding parameters for different polynucleotides can be compared in two different ways using plots such as that in Figure 3. Either we can compare the salt concentrations at which a particular value of $K\omega$ is reached, or we can extrapolate to determine $K\omega$ at a common salt concentration for each polynucleotide (assuming linearity of the log–log plots for the various polynucleotides beyond the experi-mental range; this assumption will be justified below). The former approach, at the approximate midpoint of the titration range ($K\omega \simeq 5 \times 10^5$ M^{-1}), gives salt concent-rations of 0·4 M, 0·58 M and 2·1 M for poly(rA), poly(dA) and poly(dT), respectively; the latter procedure, at salt concentrations approximately equal to physiological values ($\sim 0\cdot2$ M-NaCl; see Kao-Huang *et al.*, 1977), yields values for $K\omega$ of $3\cdot2 \times 10^7$ M^{-1}, $2\cdot0 \times 10^8$ M^{-1} and $2\cdot0 \times 10^9$ M^{-1} for the same polynucleotides.

The ultraviolet absorbance method is limited in its usefulness for such measure-ments by the small changes in hyper- or hypochromicity characteristic of most gene *32* protein–polynucleotide complexes (see Table 1). To circumvent this difficulty, and to permit accurate determinations of $K\omega$ for a larger range of polynucleotides, we used the quenching of intrinsic protein fluorescence on polynucleotide binding to establish additional values of $K\omega$.

(ii) *Salt back-titrations*

The salt "back-titration" procedure (Kowalczykowski *et al.*, 1980) was used to

determine the midpoints of the titration curves of each of several different polynucleotides at several gene *32* protein concentrations. By this means, values of L_F (and thus of $K\omega$) have been established for a number of additional poly- nucleotides as a function of salt concentration. The data obtained by this technique are plotted as log $K\omega$ *versus* log [NaCl] in Figure 4, and summarized, together with the ultraviolet titration results, in Table 2.
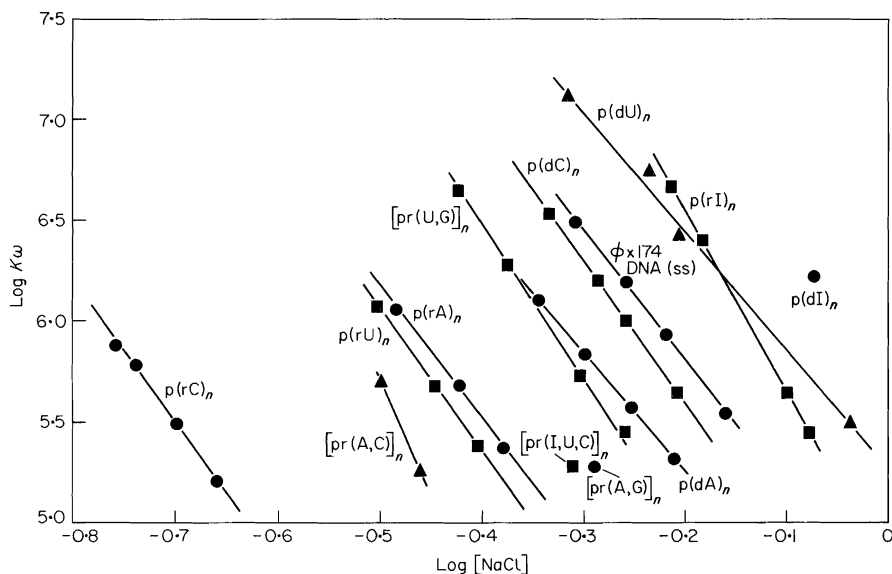

FIG. 4. Absolute values of $K\omega$ for gene *32* protein binding to various polynucleotides, determined using the fluorescence salt back-titration procedure. Buffers and conditions are as described in the legend to Table 2 (and by Kowalczykowski *et al.*, 1980; Figs 4 and 5 and Materials and Methods). Poly[r(A,C)] contained 0·83 mol A/mol total nucleotide; poly[r(U,G)] contained ~0·5 mol U/mol total nucleotide; poly[r(A,G)] contained ~0·5 mol A/mol total nucleotide; poly[r(I,U,C)] contained ~0·33 mol *each* nucleotide/mol total nucleotide.

A number of inferences can be drawn from these measurements. First, all the log $K\omega$ *versus* log [NaCl] plots appear to be linear, with slopes $\simeq -7$ ($\pm 1\cdot 5$)[†]. This linearity and essential constancy of slope for all the polynucleotides tested suggests that the conclusions drawn by Kowalczykowski *et al.* (1980) on the molecular basis of the salt dependence of co-operative binding in the polynucleotide mode apply to all polynucleotides. Furthermore, the essential constancy of the slopes of these log–log plots for different polynucleotides, over more than an order of magnitude change in NaCl concentration, strongly support the validity of the linear extrapo- lation of these plots beyond the range of salt concentrations attainable with a

[†] We note that the slopes of the poly(rI) and poly(dT) data fall outside this range. We believe that the slope of the log–log plot for the former is artificially increased by competitive double-helix formation under these conditions. At the very high salt concentrations at which the poly(dT) data were measured ([NaCl] > 1·5 M), the approximation that salt concentration $\simeq$ salt activity is clearly no longer ap- propriate. The breakdown of this approximation, plus possible changes in protein and nucleic acid hydration at these salt concentrations, may account for the anomalously low slope of the poly(dT) complex (Table 2).

TABLE 2

Comparison of binding parameters for gene 32 protein to various polynucleotides

Polynucleotide	$(\partial \log K\omega / \partial \log [\text{NaCl}])^e$	$K\omega \ (\text{M}^{-1})^f$ (at 0·2 M-NaCl)	Concn of NaCl $(\text{M})^f$ (at $K\omega \simeq 5 \times 10^5 \ \text{M}^{-1}$)
Poly(rC)	−7·0	$3·2 \times 10^5$	0·19
Poly(rU)	−7·5	$4·0 \times 10^7$	0·36
Poly(rA)	−6·5	$3·2 \times 10^7$	0·37
Poly(rI)[a]	−9·0	—	—
Poly(dC)	−7·1	$1·6 \times 10^9$	0·61
Poly(dU)	−5·5	$5·0 \times 10^8$	0·84
Poly(dA)	−5·9	$2·0 \times 10^8$	0·53
Poly(dI)	—	—	0·79
Poly[r(A,C)][b, c]	−9·0	—	0·32
Poly[r(U,G)][c]	−7·6	—	0·50
ϕX174(ss)DNA	−6·3	$1·0 \times 10^9$	0·65
Poly(rA)[b]	−7·0	$7·9 \times 10^7$	0·42
Poly(dA)[b]	−5·7	$2·0 \times 10^8$	0·58
Poly(dT)[b]	−3·5	—	2·1
Poly(rϵA)[d]	−6·5	$1·2 \times 10^9$	0·69

All measurements (unless otherwise indicated) were made by the fluorescence quenching salt back-titration procedure at 25°C in buffer C at total NaCl concentrations as indicated (see the text). Polynucleotide concentrations ranged from $1·2 \times 10^{-6}$ M to $1·5 \times 10^{-4}$ M; gene *32* protein concentrations ranged from $1·6 \times 10^{-7}$ M to 2×10^{-5} M.

[a] Forms double-stranded structures at these salt concentrations.

[b] Measured by u.v. absorbance titrations in buffer B at 23°C. Protein concentrations ranged up to $1·9 \times 10^{-5}$ M, and polynucleotide concentrations were $2·5 \times 10^{-5}$ M.

[c] Poly[r(A,C)] contained 0·83 mol A/mol total nucleotide; and poly[r(U,G)] contained 0·50 mol U/mol total nucleotide.

[d] The poly(rϵA) binding parameters are from Kowalczykowski *et al.* (1980).

[e] The standard error in these slopes is $\simeq \pm 1$.

[f] The errors in these columns will, of course, depend on the lengths of the relevant extrapolations. In general, they should not exceed $\pm 20\%$ of the listed values.

ss, single-stranded.

particular polynucleotide. Such linear extrapolations were used to obtain comparable values of $K\omega$ for different polynucleotides at a common (0·2 M-NaCl) salt concentration (Table 2).

In addition, the results illustrated by Figure 4 show unequivocally that the binding affinity of various polynucleotides does show specificity (for gene *32* protein binding in the co-operative polynucleotide binding mode). A number of generalizations emerge, which are qualitatively apparent in Figure 4 and are documented quantitatively in Table 2. (1) In each case the polydeoxyribonucleotide binds gene *32* protein more tightly than does its polyribonucleotide homologue. Differences in measured values of $K\omega$ range from $\sim 10^1$ for poly(dU) and poly(rU), to $\sim 10^4$ for poly(dC) and poly(rC). (2) There is no obvious pattern to the relative affinities for homopolynucleotides containing different bases; i.e. no systematic differences are apparent in binding affinity on the basis of purines *versus* pyrimidines, highly stacked *versus* unstacked bases, etc. (3) The net affinity of co-operatively bound

gene *32* protein for heteropolynucleotides appears to be linearly dependent on base composition; i.e. the values of $K\omega$ for poly[r(A,G)], poly[r(A,C)], poly[r(U,G)] and poly[r(I,U,C)] all fall approximately at the values expected on the basis of compositionally weighted averages of the homopolynucleotide data. Thus the measured $K\omega$ value for a given heteropolynucleotide may be calculated as:

$$(K\omega)_{obs} = \sum_i f_i(K\omega)_i, \tag{1}$$

where f_i is the fraction of each base present in the heteropolynucleotide, and $(K\omega)_i$ is the net binding affinity to the homopolynucleotide containing that base. Note that poly(rG) and poly(dG) are not listed in Table 2. As is well-known, these moieties are very prone to form "self-structures" when present as homopolynucleotides and, in our hands, yielded erratic values of $K\omega$ and slopes in plots such as Figure 4. The $K\omega$ values obtained with poly(rI) and poly(dI) appear to be more representative of the behavior of isolated ribo- or deoxyriboguanine residues as they occur in random polynucleotide copolymers or in DNA, and are so used in these calculations. In these terms we find that by substituting dI for dG and dU for dT (see below), and using the known base composition of ϕX174 DNA, we can calculate a value of $K\omega$ for this single-stranded DNA with equation (1), which is the same, within experimental error, as that measured directly for this material. (4) Poly(dT) is rather special, in that it is bound much more tightly by gene *32* protein than is any other polynucleotide tested. If we use the poly(dT) value of $K\omega$ in calculating $K\omega$ of single-stranded ϕX174 DNA by equation (1), we find that the calculated parameter exceeds by at least tenfold the value determined experimentally. This suggests that an isolated dT residue contributes to the overall affinity of gene *32* protein for the polynucleotide lattice to about the same extent as a dU residue, with two or more dT residues in sequence being required to demonstrate the increased affinity characteristic of poly(dT). This interpretation is strengthened by the observation that $K\omega$ measured for the alternating polynucleotide poly[d(A-T)] (by melting profile depression methods; see Jensen *et al.*, 1976) falls at the value expected from equation (1) using the poly(dU), rather than the poly(dT) value of $K\omega$ (Newport, unpublished data).

A quantitative summary of the data of Figure 4, in which we compare values of $\partial \log K\omega / \partial \log [NaCl]$, of $K\omega$ at constant (~ 0.2 M) salt concentration, and of the concentrations of NaCl required to make $K\omega \simeq 5 \times 10^5$ M^{-1} for each polynucleotide, is assembled in Table 2. For comparison we also list representative parameters derived from the ultraviolet light absorbance titrations (above) for some of the same polynucleotides. In general, the values of $K\omega$ obtained by the two methods fall within approximately a factor of two. Considering the very different natures of the two techniques, the different levels of protein concentrations involved, and the errors inherent in each, this represents good agreement. In addition, binding parameters obtained by enhancement of polynucleotide fluorescence titration procedures for poly(rϵA) in the preceding paper are also included in Table 2. We see that this modification enhances the net affinity of gene *32* protein ~ 30-fold above the value measured with poly(rA), while affecting the salt dependence of the binding very little.

(iii) *Binding specificity and salt dependence of* K *and* ω

In the preceding paper we showed that we could determine "best-fit" theoretical titration isotherms for gene *32* protein binding to (e.g.) poly(rεA) (Fig. 8 of Kowalczykowski *et al.*, 1980), and thus obtain separate best fit values of K and ω. By this means, as well as by measurements of K for single-stranded DNA under conditions of low protein binding density, we were able to show that virtually the entire salt dependence of gene *32* protein binding co-operatively to polynucleotides can be attributed to K, and that ω is essentially independent of salt concentration. Here we have used the same procedure to analyze binding isotherms at several salt concentrations for several different polynucleotides in order to ask whether this conclusion is general for other polynucleotides, and whether the binding specificity (differences in $K\omega$ between polynucleotides of different base and sugar composition) lies in K or in ω (or possibly in both).

To approach these questions we analyzed binding isotherms obtained by titrating poly(rA), poly(dA) and poly(rεA); the results are summarized in Table 3. As before, we observe that the experimental titration curves are not perfectly symmetrical. This effect has been largely attributed to the finite length of the polynucleotide

TABLE 3

Summary of computer "fits" to gene 32 *protein polynucleotide titration curves*

Poly-nucleotide	[NaCl] (M)	$K\omega$ ($M^{-1} \times 10^{-6}$)	ω (range) $\times 10^{-3}$	Best fit value of ω ($\times 10^{-3}$)	K (M^{-1})
Poly(rεA)	0·35	34·8	1·5– 7	3·5	$1·0 \times 10^4$
	0·35	45·0	3·0–10	5·0	$9·0 \times 10^3$
	0·40	13·8	2·5– 4·5	4·0	$3·5 \times 10^3$
	0·45	7·04	4·5–10	5·0	$1·4 \times 10^3$
	0·50	4·20	1·5– 4·5	3·5	$1·2 \times 10^3$
	0·60	1·24	1·5– 4·5	3·5	$3·5 \times 10^2$
Poly(dA)	0·49	1·22	2·0– 7·0	6·0	$2·0 \times 10^2$
	0·59	0·51	3·0–10	6·0	$8·5 \times 10^1$
	0·78	0·10	2·5– 5·0	6·0	$2·9 \times 10^1$
Poly(rA)	0·35	1·69	1·5– 4·0	2·0	$8·5 \times 10^2$
	0·40	0·82	1·5– 4·5	2·0	$4·1 \times 10^2$
	0·45	0·25	1·5– 2·5	2·5	$1·0 \times 10^2$
	0·50	0·14	1·0– 2·0	2·5	$0·7 \times 10^2$

Titrations were carried out, and the data analyzed, as described for Figs 3 and 4 and Table 2, and as described by Kowalczykowski *et al.* (1980).

lattices used in these titrations (Kowalczykowski *et al.*, 1980); this interpretation was confirmed by fractionating poly(rεA) on a Sepharose CL-4B column, and using only the largest fraction (void volume) in a replicate titration with gene *32* protein (unpublished data). As expected, the top "break" of the isotherm was appreciably "less rounded" than that obtained with the unfractionated sample of poly(rεA), while the shape of the first (bottom) break was essentially unaffected. For this

reason the best fit values listed in Table 3 reflect primarily the analysis of the lower 50% of the binding isotherm.

We tentatively conclude from the data of Table 3 that the average values of ω obtained are essentially salt independent for all the polynucleotides examined, and that the binding specificity differences are *largely* in K. (However, some difference may exist in ω in certain cases; e.g. Table 3 suggests that ω may be ~ 3-fold larger for poly(dA) than for poly(rA).)

4. Discussion

In this paper we have shown that gene *32* protein, complexed with nucleic acid lattices in the polynucleotide binding mode, shows significant differential affinity for polynucleotides of differing base composition and sugar type. Here we consider first the possible molecular bases of these affinity differences, and then show how these differences might provide a quantitative explanation of the *biological* specificities inherent in the function of this protein in DNA replication and in the control of its own synthesis.

(a) *Molecular aspects of binding specificity*

As summarized above (Fig. 4 and Table 2), gene *32* protein discriminates much more effectively between different nucleic acids when binding in the polynucleotide binding mode than in the oligonucleotide mode. In the polynucleotide mode, the net binding affinity of the protein for homopolynucleotides of differing base and sugar composition can differ by factors as large as 10^4 in $K\omega$. As shown in detail in this paper, these specificities in net binding affinity (except for poly(dT)) seem to depend on differences in the affinity of the protein for individual nucleotide residues. In terms of Figure 13 of Kowalczykowski *et al.* (1980), we might attribute these differences to small conformational changes in the left-hand (XpX) binding sub-site of the protein, resulting in additional contacts with the functional groups of the bases in the polynucleotide binding mode, and thus in increased binding specificity. We have also speculated in the preceding paper that some part of the increased binding affinity for DNA over RNA chains may be due to the greater ease with which the backbone of the former can be deformed (Sundaralingam, 1975) to accommodate somewhat "out-of-register" spacings between binding sub-sites on the protein surface. In addition, in some cases (particularly for poly(dT)), it appears that at moderate salt concentrations the intrinsic binding constant (K) of the protein in the polynucleotide mode exceeds that in the oligonucleotide mode; why then does the protein not bind to (e.g.) (dT)$_8$ in the former mode? We speculate that binding of the protein to oligonucleotides of length (l) ≤ 8 residues in the polynucleotide mode may result in specifically unfavorable interactions; e.g. the terminal residues of oligonucleotides carry less bound counterions, and thus engage in weaker electrostatic interactions, than "interior" polynucleotide residues (see Kowalczykowski *et al.*, 1980). Also the terminal residues of oligonucleotides have more conformational freedom than interior polynucleotide residues, again weakening interactions relatively more because more conformational entropy is lost on binding. In any case, the protein seems to bind to short ($l = 2$ to 8 residues)

oligonucleotides in the oligonucleotide binding conformation under all salt concentration conditions tested. Final molecular explanations of these and other binding affinity differences must probably await X-ray crystallographic elucidation of the molecular structures of the protein in both its nucleic acid binding conformations.

(b) *Biological specificity and the control of gene 32-protein function*

Control of genome expression *via* protein–nucleic acid interactions requires binding specificity. At the simplest level it requires that a particular stretch of nucleic acid be complexed with protein in preference to all the competing nucleic acid lattices that are also present in the cell. Nature has solved the problem of developing binding specificity in a number of ways. For recognition of a particular sequence of double-stranded nucleic acid base-pairs, as in repressor–operator interactions, the specificity inherent in the polyfunctional array of hydrogen bond donors and acceptors present in the grooves of the double-helical region has been exploited by the development of proteins carrying complementary arrays of hydrogen-bonding groups (see von Hippel, 1979). Helix destabilizing proteins, such as T4-coded gene *32* protein, recognize target lattices primarily on the basis of *strandedness* rather than base sequence; that is, they bind preferentially and specifically to single-stranded regions (present within the predominantly double-stranded genome) which arise as intermediates in the processes of replication, recombination, etc. Specificity discrimination here is only between single and double-stranded structures; the actual *sequence* of nucleotide residues appears to be of relatively little importance.

Binding co-operativity strengthens this apparent lack of dependence on nucleotide sequence, since it ensures that contiguous binding of protein will be preferred over isolated binding under almost all conditions. Figure 4 and Table 2 show that only for exceptional sequences (e.g. a long run of dT residues) is the difference in $K\omega$ between one sequence and another likely to exceed the additional free energy cost inherent in initiating a new protein cluster *versus* extending a pre-formed one. In terms of the function of the protein in replication and recombination, co-operativity is also required to bring about *complete* coverage of single-stranded sequences (thus protecting the lattice against single-strand specific nucleases), and to facilitate the removal (by competitive, with double helix formation, binding of gene *32* protein) of the small and relatively unstable hairpin structures that can form adventitiously within short, single-stranded DNA sequences (see McGhee & von Hippel, 1974; von Hippel *et al.*, 1977).

The genetic studies reported by Krisch *et al.* (1974) and Russell *et al.* (1976), together with the biochemical findings of Lemaire *et al.* (1978), suggested that the control of gene *32* protein synthesis depends on the existence of an effective sequential binding specificity: first the available single-stranded DNA sequences in the cell must be titrated with protein and then, after the attainment of a threshold protein concentration, binding to some critical site(s) on the gene *32* protein message (which reversibly prevents its further use in translation) must follow. The net binding affinity of the protein for the target site(s) on gene *32* protein mRNA must be greater than that to the control (initiation?) sites on other T4 mRNAs, since an

approximately threefold excess (Lemaire *et al.*, 1978) of free gene *32* protein is required before translation of these mRNAs is also inhibited in a mixed *in vitro* protein synthesis system. This type of control of binding specificity, based on autogenous regulation of the synthesis of a limited amount of binding protein, and its sequential distribution between binding "sinks" (lattices) of differing affinities, finds ready explanation in the co-operativity of binding of gene *32* protein.

(c) *Co-operativity and binding specificity*

We may see that, in principle, the actual *differences* in binding affinity of the protein for the various nucleic acid lattices that offer competing binding sites need not be great to lead to this result. If, for example, the difference in K for hypothetical lattices X and Y were a factor of two (favoring binding to poly(X)), then under conditions of equal (excess) concentration of both lattice types the equilibrium probability that the first protein molecule added would bind to poly(X) is just twice the probability that it would bind to poly(Y), clearly a low level of specificity. On the other hand, because of co-operativity (we assume in this example that ω is the *same* for protein binding to poly(X) and poly(Y)), the proteins added will tend to bind in co-operative (contiguous) clusters. Thus, if binding at equilibrium involves average clusters that are *two* proteins in length, the distribution ratio for a two protein-cluster binding competitively to poly(X) and poly(Y) would be $\sim 2^2 : 1$. If the average cluster size were c, the preference for poly(X) over poly(Y) in this example would be $2^c : 1$. Since c would be ~ 12 for a lattice one-half saturated with protein with $\omega = 10^3$ (McGhee & von Hippel, 1974), the net preference for poly(X) over poly(Y) in this hypothetical example would be 4096:1. This represents a very high level of specificity indeed, which we note is actually attained *via* amplification (by co-operativity) of an intrinsic specificity, in terms of monomer protein affinity, of only 2:1 in favor of poly(X).

The competition experiment shown in Figure 2, in which poly(rA) competes with a tenfold excess of poly[r(A,C)] for a limited amount of gene *32* protein, illustrates this principle quantitatively. This particular pair of polynucleotides was chosen for the experiment because their net binding affinities for gene *32* protein differ by exactly tenfold in $K\omega$ (favoring poly(rA); see u.v. titration data, Table 2). A tenfold excess of poly[r(A,C)] over poly(rA) was therefore placed in the cuvette, and a u.v. absorbance titration performed. Figure 2 shows that *all* the poly(rA) is titrated to completion before titration of the poly[r(A,C)] begins. Clearly these results are in total accord with the model of specificity based on competitive co-operativity of binding presented above.

(d) *Autogenous regulation of gene* 32 *protein synthesis*

The results of this paper, together with the enhancement of intrinsic binding specificity by co-operativity demonstrated above, can be used to account quantitatively for the sequential specificity of gene *32* protein binding which autoregulates the synthesis of the protein. Thus the fact that gene *32* protein will first bind intracellularly to single-stranded DNA sequences is guaranteed by the preference of the protein for deoxyribose over ribose-containing nucleic acids. In addition, the

preference for clusters of dT residues over all others will also favor binding to DNA over RNA sequences. After all DNA sequences are titrated, the concentration of free protein builds up by further synthesis until a free gene *32* protein concentration adequate to reach the gene *32* mRNA binding threshold is attained. Binding then continues until the critical control region(s) of gene *32* mRNA is (are) saturated. At this point further synthesis is (reversibly) suspended.

This system requires that the control site of gene *32* mRNA saturates *before* those of the other T4 mRNAs. This could occur either because this sequence is especially rich in (e.g.) rG residues or, in keeping with the original model of Russel *et al.* (1976), because the initiation site for ribosome binding in this mRNA comprises the *longest* such sequence in this family of mRNAs that is *not* encumbered with double-stranded hairpin loops too stable to be melted out by gene *32* protein under *in vivo* conditions†. Qualitatively, one can see why longer sequences (all other aspects of binding affinity being equal) will bind gene *32* protein first; essentially one "unit" of ω (~ 4 kcal/mol for gene *32* protein) in binding free energy is "lost" every time a new binding protein cluster is initiated, rather than continuing, by contiguous binding, a pre-existing cluster.

Clearly, small differences in binding affinity (K), co-operativity (ω) and lattice length (l) between competing binding lattices can greatly perturb the distribution of a limited amount of protein over such a set of lattices, and thus can also greatly perturb the apparent binding specificity. Elsewhere we will present a series of model calculations to simulate quantitatively the possibilities for genome control inherent in competitive co-operative protein binding. To provide a first approximation model of the gene *32* protein control system, and to illustrate the quantitative possibilities inherent even in a simple manipulation of the *lengths* of the competing nucleic acid lattices, we present a possibly relevant model calculation in Figure 5. Here we use a value of K corresponding to that expected for DNA sequences with the base composition characteristic of whole T4 DNA in an ionic environment containing $\sim 0 \cdot 3$ M-NaCl (comparable to that used in the *in vitro* repression studies reported by Lemaire *et al.* (1978)). Even in this simple system we see that, after a short lag required to build up a threshold concentration of free gene *32* protein in the solution, single-stranded DNA sequences (taken for this example to be greater than 120 nucleotide residues in effective length) are complexed completely ($> 99\%$). After a further lag, now to permit the free gene *32* protein concentration to rise to a second threshold level, gene *32* mRNA control regions are complexed. Finally, after an additional threefold increase in L_F, the control regions of the other T4 mRNA molecules are saturated. In this model system we have used a lattice length of ~ 50 nucleotide residues of average composition for the open (non-hairpin loop-containing control) regions(s) of gene *32* mRNA, and an average lattice length of ~ 29 nucleotide residues of the same average composition for the control sequences of the other T4 mRNAs. (The use of different values of K for DNA and for the different mRNA lattices could obviously enhance specificity further.) The theoretical titration curves for the DNA sequences are close enough to the infinite lattice

† Some mRNA sequences not involved in control could bind gene *32* protein at lower free protein concentrations than the control sequence(s); it is only required that proteins bound in these regions not interfere with the elongation (as opposed to the initiation) phase of mRNA translation.
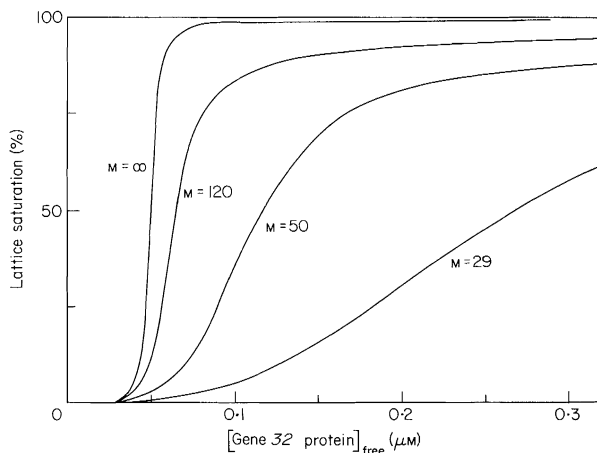
limit to permit the use of equation (15) of McGhee & von Hippel (1974); for the mRNA lattices the finite lattice binding equations of Epstein (1979) have been used. Other details of the calculation are given in the Figure legend.

Obviously the development of a *specific* quantitative model for this system will require actual information on the secondary structure and base composition of the initiation sequences of the relevant mRNAs (see Krisch *et al.*, 1980). However, we emphasize that the system is not highly dependent on the exact values of lattice length or binding parameters used. Thus, even before the necessary structural data on the T4 mRNAs are at hand, we can see that systems of this *type* are entirely capable of providing the hierarchy of binding specificity required for the control of gene *32* protein synthesis and function. Clearly such model calculations also provide clues as to the sorts of compositional and structural properties of mRNA that might be relevant in such a translational control system.

REFERENCES

Bobst, A. M. & Pan, Y.-C. E. (1975). *Biochem. Biophys. Res. Commun.* **67**, 562–570.
Epstein, I. R. (1979). *Biophys. Chem.* **8**, 327–339.
Jensen, D., Kelly, R. C. & von Hippel, P. H. (1976). *J. Biol. Chem.* **251**, 7215–7228.

Kao-Huang, Y., Revzin, A., Butler, A. P., O'Connor, P., Noble, D. W. & von Hippel, P. H. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 4228–4232.

Kelly, R. C., Jensen, D. E. & von Hippel, P. H. (1976). *J. Biol. Chem.* **251**, 7240–7250.

Kowalczykowski, S. C., Lonberg, N., Newport, J. W. & von Hippel, P. H. (1980). *J. Mol. Biol.* **145**, 75–104.

Krisch, H. M., Bolle, A. & Epstein, R. H. (1974). *J. Mol. Biol.* **88**, 89–104.

Krisch, H. M., Allet, B. & Duvoisin, R. M. (1980). In *Mechanistic Studies of DNA Replication and Genetic Recombination, ICN-UCLA Symposium on Molecular and Cellular Biology* (Alberts, B. & Fox, C. F., eds), vol. 19, Academic Press, New York.

Lemaire, G., Gold, L. & Yarus, M. (1978). *J. Mol. Biol.* **126**, 73–90.

Lonberg, N., Kowalczykowski, S. C., Paul, L. S. & von Hippel, P. H. (1980). *J. Mol. Biol.* **145**, 123–138.

McGhee, J. D. & von Hippel, P. H. (1974). *J. Mol. Biol.* **86**, 469–489.

McGhee, J. D. & von Hippel, P. H. (1977). *Biochemistry*, **16**, 3267–3276.

Russel, M., Gold, L., Morrissett, H. & O'Farrell, P. Z. (1976). *J. Biol. Chem.* **251**, 7263–7270.

Sundaralingam, M. (1975). In *Structure and Conformation of Nucleic Acids and Protein Nucleic–Acid Interactions* (Sundaralingam, M. & Rao, S. T., eds), pp. 487–524, University Park Press, Baltimore.

von Hippel, P. H. (1979). In *Biological Regulation and Development* (Goldberger, R., ed.), vol. 1, pp. 279–347, Plenum Publishing Corporation, New York.

von Hippel, P. H., Jenson, D. E., Kelly, R. C. & McGhee, J. G. (1977). In *Nucleic Acid–Protein Recognition* (Vogel, H. J., ed.), pp. 65–89, Academic Press, New York.