Balanced Branching in Transcription Termination

K. J. Harrington and R. B. Laughlin
Department of Physics, Stanford University, Stanford, California 94305

S. Liang
NASA Ames Research Center, Moffett Field, California 94035
(April 27, 2001)

The theory of stochastic transcription termination based on free-energy competition[1] requires two or more reaction rates to be delicately balanced over a wide range of physical conditions. A large body of work on glasses and large molecules suggests that this should be impossible in such a large system in the absence of a new organizing principle of matter. We review the experimental literature of termination and find no evidence for such a principle but many troubling inconsistencies, most notably anomalous memory effects. These suggest that termination has a deterministic component and may conceivably be not stochastic at all. We find that a key experiment by Wilson and von Hippel[2] allegedly refuting deterministic termination was an incorrectly analyzed regulatory effect of Mg^{2+} binding.

PACS numbers: 87.15.-v, 82.20.Pm, 61.43.Fs

I. INTRODUCTION

The branching ratio of the termination process in gene transcription is balanced. In the case most thoroughly studied, ρ-independent termination in procaryotes, conventional gel experiments performed *in vitro* find a fraction P of elongating RNA polymerase reading through the termination sequence with $|\ln(1/P-1)|<4$ essentially always, even though P is different for different terminators and can be made to exhibit order-1 changes by perturbing the environment. This effect is astonishing from the standpoint of microscopic physics because a stochastic decision to read through or not requires a competition of transition rates - quantities of inverse time - that must be nearly equal for the branching to be balanced. RNA polymerase, however, is more the size of a glass simulation than a small molecule and thus possesses a broad spectrum of natural time scales spanning many decades. Without some physical reason for a particular scale to be preferred, rate competition ought to have been severely unbalanced, meaning that one event occurs essentially always and the other never. Balanced branching in termination has been implicated in transcription regulation in a few cases,[3] but its broader significance, especially its robustness, is still a mystery.

In this paper we examine the experimental facts relevant to the physical nature of termination with the goal of determining what, if any, principle selects the time scale for stochastic rate balance. Our conclusion is both surprising and unsettling. We find no evidence for such a principle, but glaring weaknesses in the case for stochasticity and a large body of unexplained experimental results pointing to a termination decision that is partially deterministic. In light of the inaccessability of systems this large to ab-initio computation we conclude that transcription termination is a fundamentally unsolved problem in mesoscopic physics and an ideal target for the emerging techniques of nanoscience.

II. TERMINATION EFFICIENCY

The simplest termination sequences are the ρ-independent terminators of procaryotes, which are capable of causing polymerase to terminate *in vitro* in the absence of the ρ protein factor. A representative sampling of these is reproduced in Table I. This differs from lists that have appeared in the literature before[4,5] by having been rechecked against the fully-sequenced genome[6] and expunged of "theoretical" terminators identified only by computer search. They conform for the most part to the motif of a palindrome of typically 10 base pairs followed by a short poly-T stretch, although there is tremendous variety in the length and composition of the palindrome, variation in the length of the poly-T stretch, and occasional extension of the palindrome to include the poly-T stretch. This enormous variability contrasts with the simplicity of stop codons, which terminate protein synthesis by ribosomes and have no other function.

ρ-independent terminators are characterized by "efficiencies", i.e., the fraction of assayed transcripts that terminate. These rarely take on extreme values close to 1 or 0 when measured *in vitro*. In cases where a measurement *in vivo* exists as well the latter is usually larger[7] and is occasionally unmeasurably close to 1. Balanced termination efficiency is commonly observed *in vivo* as well, however. Table II shows results from a particularly careful study[7] *in vitro* in which termination probabilities in *E. coli* for wild-type terminators, mutant terminators, phage terminators,[8] and terminators from *S. Boydii* were measured under identical conditions. Despite the great variety of these sequences the termination efficiency runs only

Sequence[4,5]	Name	Address[6]	±	Reference
CGTTAATCCGCAAATAACGTAAAAACCCGCTTCGGCGGGTTTTTTTATGGGGGGA	rpoC t	4187152	+	RNA polymerase operon[57]
CAGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAGG	M1-RNA	3267812	-	RNA of RNase P[58]
CGTACCCCAGCCACATTAAAAAAGCTCGCTTCGGCGAGCTTTTTGCTTTTCTGCG	sup	0695610	-	supBC tRNA operon[59]
ACACTAATCGAACCCGGCTCAAAGACCCGCTGCGGCGGGTTTTTTTGTCTGTAAT		1260102	-	Nucleotide synthesis[60]
AGTAATCTGAAGCAACGTAAAAAAACCCGCCCCGGCGGGTTTTTTTATACCCGTA	L17	3437202	-	Ribosomal RNA operon[61]
TCTCGCTTTGATGTAACAAAAAACCCCGCCCCGGCGGGGTTTTTTGTTATCTGCT	rpm	3808820	-	Ribosome rpm operon[62]
GAGTAAGGTTGCCATTTGCCCTCCGCTGCGGCGGGGGGCTTTTAACCGGGCAGGA	t2	3306624	-	Polynucleotide phosphorylase[63]
CGATTGCCTTGTGAAGCCGGAGCGGGAGACTGCTCCGGCTTTTTAGTATCTATTC	deo t	4619189	+	deo operon[64]
CGTAAAGAAATCAGATACCCGCCCGCCTAATGAGCGGGCTTTTTTTTGAACAAAA	trp a	1321015	-	tryptophan synthesis[65]
GCGCAGTTAATCCCACAGCCGCCAGTTCCGCTGGCGGCATTTTAACTTTCTTTAA	trp t	1314395	-	tryptophan synthesis[66]
AAATCAGGCTGATGGCTGGTGACTTTTTAGTCACCAGCCTTTTTGCGCTGTAAGG	rplL t	4178530	+	Ribosomal proteins L7/L12[67]
AGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAA	thr a	0000263	+	threonine operon[68]
AGCACGCAGTCAAACAAAAAACCCGCGCCATTGCGCGGGTTTTTTTATGCCCGAA	leu a	0083564	-	leucine synthesis[69]
CCCGTTGATCACCCATTCCCAGCCCCTCAATCGAGGGGCTTTTTTTTGCCCAGGC	pyrBI a	4469985	-	pyrimidine synthesis[70]
ACACGATTCCAAAACCCCGCCGGCGCAAACCGGGCGGGGTTTTTCGTTTAAGCAC	ilvB a	3850449	-	ilvB operon[71]
GAAACGGAAAACAGCGCCTGAAAGCCTCCCAGTGGAGGCTTTTTTTGTATGCGCG	pheS a	1797160	-	Phenylalanyl-tRNA synthetase[72]
CTTAACGAACTAAGACCCCCGCACCGAAAGGTCCGGGGGTTTTTTTTGACCTTAA	ilvGEDA a	3948053	+	ilvGEDA operon[73]
CCGCCCCTGCCAGAAATCATCCTTAGCGAAACGTAAGGATTTTTTTTATCTGAAA	rrnC t	3944645	+	Ribosomal RNA operon[74]
CATCAAATAAAACAAAAGGCTCAGTCGGAAGACTGGGCCTTTTGTTTTATCTGTT	rrnD t	3421006	+	Ribosomal RNA operon[75]
TCCGCCACTTATTAAGAAGCCTCGAGTTAACGCTCGAGGTTTTTTTTCGTCTGTA	rrnF (G) t	0228998	+	Ribosomal RNA operon[76]
GCATCGCCAATGTAAATCCGGCCCGCCTATGGCGGGGCCGTTTTGTATGGAAACCA	frdB t	4376529	-	Fumarate reductase[77]
TGAATATTTTAGCCGCCCCAGTCAGTAATGACTGGGGCGTTTTTTATTGGGCGAA	spot42-RNA	4047542	+	spot42 RNA[78]
ATTCAGTAAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTATTACTCA	tonB t	1309824	+	Membrane protein[79]
AGAAACAGCAAACAATCCAAAACGCCGCGTTCAGCGGCGTTTTTTCTGCTTTTCT	glnS T	0707159	+	Glutaminyl-tRNA synthetase[80]
CTGGCATAAGCCAGTTGAAAGAGGGAGCTAGTCTCCCTCTTTTCGTTTCAACGCC	rplT t	1797371	-	Ribosome protein L20[81]
GCATCGCCAATGTAAATCCGGCCCGCCTATGGCGGGCCGTTTTGTATGGAAACCA	ampC a	4376529	-	β-lactamase[82]
TGCGAAGACGAACAATAAGGCCTCCCAAATCGGGGGGCCTTTTTTATTGATAACA	phe a	2735697	+	Phenylalanine operon[83]
ACGCATGAGAAAGCCCCCGGAAGATCACCTTCCGGGGGCTTTTTTATTGCGCGGT	hisG a	2088121	+	ATP synthesis[84]
CATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTT	rrnB t_1	4169333	+	Ribosomal RNA operon[85,86]
GGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACA	rrnB t_2	4169493	+	Ribosomal RNA operon[85,86]
AATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTT	lacI tII	0365588	-	Lactose synthesis[87]
CTTTTTGGCGGAGGGCGTTGCGCTTCTCCGCCCAACCTATTTTTACGCGGCGGTG	uvrD a	3995538	+	DNA helicase II[88]

TABLE I. ρ-independent terminators in *E. coli* taken primarily from Brendel et al.[4] These are oriented in the reading direction and are aligned at the poly-T stretch. The palindrome is underlined. The beginning and end of the selected sequences have no absolute meaning but simply follow the convention of d'Aubenton et al.[5] The address identifies the location in the standard *E. coli* genome[6] of the left-most nucleotide in the table.

Sequence	Name	% T
GGCTCAGTCGAAAGACTGGGCCTTTCGTTTTAAT	rrnB t_1	84 ± 1
TCAAAAGCCTCCGACCGGAGGCTTTTGACTATTA	tonB t	19 ± 1
CCAGCCCGCCTAATGAGCGGGCTTTTTTTTGAAC	trp a	71 ± 2
CCAGCCCGCCTAATGAGCGGGCTTTTGCAAGGTT	trp a 1419	2 ± 1
CCAGCCCGCCTAATAAGCGGGCTTTTTTTTGAAC	trp a L126	65 ± 4
CCAGCCCGCCTAATAAGCGGACTTTTTTTTGAAC	trp a L153	8 ± 4
CTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTA	$T7T_e$	88 ± 2
GGCTCACCTTCACGGGTGAGCCTTTCTTCGTTCX	$T3T_e$	14 ± 2
GGCCTGCTGGTAATCGCAGGCCTTTTTATTTGGG	tR2	49 ± 4
AAACCACCGTTGGTTAGCGGTGGTTTTTTGTTTG	RNA I	73 ± 4

TABLE II. Termination efficiencies measured *in vitro*.[7] The first 3 terminators are native to *E. coli*. These are followed by 3 mutants, 3 phage terminators,[8] and one from *S. Boydii*. Far-right underlined sequences are termination zones.

Sequence	Name	% T
GTTAATAACAGGCCTGCTGGTAATCGCAGGCCTTTTTATT	tR2	40
GTTAATAACAGGGGACGTGGTAATCCGTCCCCTTTTTATT	tR2-6	56
TAATAACAGGCCTGGCTGGTAATCGCCAGGCCTTTTTATT	tR2-11	54
CCGGGTTAATAACAGGCCTGCTTCGGCAGGCCTTTTTATT	tR2-12	69
CGGGTTATTAACAGGCCTCTGGTAATCGAGGCTTTTTATT	tR2-13	11
ATAACAGGGGACGTGGTAATCGCCAGCAGGCCTTTTTATT	tR2-14	20
GTTAATAAAAGGCCTGCTGGTAATCGCAGGCCTTTTTATT	tR2-16	36
GGTTCTTCTCGGCCTGCTGGTAATCGCAGGCCTTTTTATT	tR2-17	67

TABLE III. Termination efficiencies for modified versions of the phage λ terminator tR2.[11]

from 2% to 88%. Many other researchers report similar values for terminators in *E. coli* and other bacteria,[9] including artificially altered terminators.[10]

Sequence	Name	rpo+	rpo203
GCAACCGCTGGGGAATTCCCCAGTTTTCA	trpC 301	0	20
AACCGCTGGCCGGGATCGGCCAGTTTTCA	trpC 302	8	35
CAGCCGCCAGTTCCGCTGGCGGCTTTTAA	trp t	25	45
ACCAGCCCGCCTAATGAGCGGGCTTTTGC	trp a 1419	3	35
CAGCCCGCCTAATGAGCGGGCTGTTTTTT	trp a 135	65	80

TABLE IV. Termination efficiences for wild-type *E. coli* polymerase (rpo+) and mutant polymerase (rpo203).[12] *trp t* is native to the genome. The rest are either mutants or synthetic.

The results in Tables III and IV show balanced termination for modified versions of the phage terminator $tR2$[11] and for mutant polymerase.[12] This also makes order-1 changes to the efficiencies themselves. Similar effects were reported by other researchers[9,13] with different mutant polymerases. Modifications up to 20 base pairs upstream and downstream of the terminator cause large changes to the efficiency without causing it to unbalance.[7] Thus balanced termination efficiency is the norm rather than the exception.

III. LARGE MOLECULES AND GLASSES

Large systems are qualitatively different from small ones.[14] The specific heat of all non-crystalline matter in macroscopic quantities - including biological matter - is proportional to T at low temperatures.[15] This behavior is fundamentally incompatible with the linear vibration of the atoms around sites, and is caused by collective quantum tunneling of atoms between energetically equivalent "frustrated" configurations.[16] It contrasts sharply with the T^3 behavior of crystals with small unit cells. Glasses also exhibit stretched-exponential time dependence in response to perturbations, i.e., of the form $\exp(-At^\beta)$ with $\beta < 1$, indicating a broad spectrum of decay rates rather than just one. They also exhibit memory effects, such as "remanence" in spin glasses[17] or the well-known failure of ordinary silica to crystallize without annealing. This behavior is universal and robust. All non-crystalline macroscopic matter exhibits hysteresis, metastability, a broad spectrum of relaxation times, and memory.

How large a system must be before it can exhibit such behavior is not known, as the relevant experiments are difficult to perform except on macroscopic samples, but there are many indications that even medium-sized proteins have glass-like properties. Crystals of myoglobin, a protein with a molecular weight of only 17,000, have linear specific heats at low temperatures[18] and exhibit stretched-exponential response to photodissociation pulses.[19] Denatured proteins refold on a variety of time scales ranging from nanoseconds to seconds,[20] and amino acids sequences chosen at random will not fold at all.[21] Permanent misfolding of proteins with molecular weights of only 30,000 has been implicated in prion diseases.[22] Many enzymes exhibit hysteresis in their catalytic rates.[23,24] The activity of cholesterol oxidase of *Brevibacterium sp.*, a protein with molecular weight 53,000, was recently shown by fluorescence correlation techniques to have a memory effect persisting about 1 second under normal conditions at room temperature.[25] Other notable examples include wheat germ hexokinase (mol. wt. 50,000[26]) with a half-life of 2 minutes,[27] rat liver glucokinase (mol. wt. 52,000[28]) at 1 minute,[23] and yeast hexokinase (mol. wt. 50,000) at 1-2 minutes.[29] Thus RNA polymerase complexes, which have a molecular weight of 379,000 and are comparable in size to the largest computer simulations of glasses ever performed, are good candidates for systems that exhibit glassy behavior.

Glassiness in enzymes is not always easy to observe. The mnemonic effect in yeast hexokinase occurs when it is preincubated with MgATP and free Mg^{2+} and the reaction is started with glucose, or preincubated with glucose and free Mg^{2+} and started with MgATP, but *not* if the enzyme is preincubated with glucose and metal-free ATP and then started with Mg^{2+}.[23] Mnemonic behavior can be destroyed by "desensitizing" the enzyme with contaminants.[26] Time scales can depend on enzyme, substrate, product, activator and effector ligand concentrations as well as pH, buffers, and temperature.[23,29,30] Before hysteresis and memory effects were recognized, early investigators generally adjusted such reaction conditions until the "improper" behavior was eliminated.[23]

IV. POLYMERASE STATES

While the size of RNA polymerase makes it plausible to expect glassy behavior on purely theoretical grounds, several direct lines of evidence indicate that the enzyme exhibits a spectrum of multiconformational, mnemonic and hysteretic behavior:

1. Polymerase has a catalytic mode distinct from RNA synthesis, as it can cleave the RNA transcript through hydrolysis (rather than pyrophosphorolysis, the reverse reaction of RNA synthesis),[31] with the cleavage reaction requiring Mg^{2+},[31] being template-dependent,[32] changing the polymerase footprint size,[33] and stimulated either by GreA and GreB proteins[34,35] or by high pH (8.5-10.0).[36] The last effect was discovered serendipitously, going unobserved for decades because assay conditions were being optimized to maximize elongation rates, which occur at lower pH values (7.8-8.2[37]).[36]

2. RNA polymerase mobilities in non-denaturing electrophoresis gels show significant and discontinuous variance while bearing nearly identical transcripts or identical length transcripts with different sequences.[38] These mobility variances are still observed if the RNA transcript is first removed by ribonuclease digestion.[39]

3. RNA polymerase ternary complexes vary greatly in their stability and mode of binding to DNA (ionic or non-ionic) in a template-dependent manner. Some complexes are stable against very high salt concentrations ($[K^+] = 1$ M), while others (specifically those proximal to an upstream palindrome sequence) are salt-sensitive (completely dissociating in concentrations as low as 20 mM K^+). However, the salt-sensitive complexes are stabilized by millimolar concentrations of Mg^{2+}.[40]

4. The size of the RNA polymerase footprint on the DNA template measured by ribonuclease digestion is significantly altered even at adjacent template positions, suggesting that the enzyme assumes different conformations during elongation.[41]

5. Guanosine tetraphosphate (ppGpp) inhibits the rate of elongation on natural DNA templates but not on synthetic dinucleotide polymer templates, and does not inhibit elongation by competing with NTP binding, but by enhancing pausing. It must therefore bind to polymerase and modify its behavior at an unrelated regulatory site in an allosteric manner, rather than interfering with the substrate binding site.[42]

6. The stability of a stalled elongation complex depends on whether the polymerase arrives at the stall site via synthesis or pyrophosphorolysis.[43]

7. Termination efficiencies are affected by transcribed upstream sequences and *un*transcribed downstream sequences adjacent to the terminator.[44]

8. Stalling elongating polymerase complexes (via nucleotide starvation) and then restarting them by nucleotide addition perturbs pausing patterns 50-60 base pairs downstream.[45]

9. An elongating polymerase's Michaelis constants K_S for NTPs vary over 500-fold for different DNA template positions,[46] and for different templates,[47] although these effects are not observed for synthetic dinucleotide polymer templates.[47]

10. The rate of misincorporation at a single site for which the correct NTP is absent is significantly different before and after isolation of ternary complexes.[48]

11. Stalled polymerase gradually "arrests" (i.e., is incapable of elongating when supplied with NTPs), with the approximate half-time for arrest estimated at 5 minutes[40] and 10 minutes[49] for different DNA templates. The polymerase can continue elongating if reactivated by pyrophosphorolysis.[40]

12. Even after undergoing arrest, crosslinking experiments show that the internal structure of polymerase gradually changes over the course of the next hour.[49]

13. Observations of single elongating RNA polymerase molecules show that it has two elongation modes with different intrinsic transcription rates and propensities to pause and arrest.[50]

The possibility of metastability - through shape memory or the conditional attachment of factors - is directly relevant to the rate-balance conundrum because it provides a simple alternative to balanced stochastic branching that requires no physical miracles. If, for example, the polymerase possessed a small number of metastable configurational states and terminated deterministically depending on which state it was in, then balanced branching would be a simple, automatic consequence of scrambling the state populations.

V. THERMAL ACTIVATION

The idea that polymerase memory is potentially relevant to expression regulation is not new.[47] It is implicit in the work of Goliger et al[51] and Telesnitsky and Chamberlin[44] and even explicitly speculated by the latter in print. However, because of the experimental evidence supporting the stochastic model of termination[1] and the widespread belief - unjustified, in our view - that proteins equilibrate rapidly, this suggestion generated little enthusiasm. A key experiment supporting the stochastic model by Wilson and von Hippel[2] is both historically important and typical, so it is appropriate that we consider it carefully.

Wilson and von Hippel promoted and stalled RNA polymerase 8 base pairs upstream of the tR2 terminator hairpin of phage λ *in vitro*, thermally equilibrated at temperature T, and then launched it forward by adding NTP. The results are reproduced in Fig. 1a. Termination occurred at sites 7, 8, and 9 base pairs downstream of the beginning of the poly-T stretch (cf. Table II) with probabilities $P_7 = N_7/N$, $P_8 = N_8/N$ and $P_9 = N_9/N$. The data were originally reported as a semilogarithmic plot of $1/\hat{P} - 1$ against temperature, where $\hat{P}_7 = N_7/N$, $\hat{P}_8 = N_8/(N - N_7)$ and $\hat{P}_9 = N_9/(N - N_7 - N_8)$. They concluded that all three branching probabilities $\hat{P}$ were thermally activated and had distinctly different activation energies. However, it is clear from Fig. 1a that this conclusion is false. The three probabilities P are essentially the same function and are well characterized by the sum $P = P_7 + P_8 + P_9$, also plotted in Fig. 1a. This is shown more explicitly in Fig. 1b, where the ratios P_7/P, P_8/P, and P_9/P are plotted against temperature. The flatness of these curves shows that the branching ratios among the three sites are essentially constant and independent of temperature within the error bars of the experiment. Note that these fractions are also all of order 1. Thus the alleged spread in activation energies was an artifact of the plotting procedure.

Let us now consider the temperature dependence. It may be seen from Fig. 1a that P saturates to 1 at 80 °C, the temperature at which Wilson and von Hippel

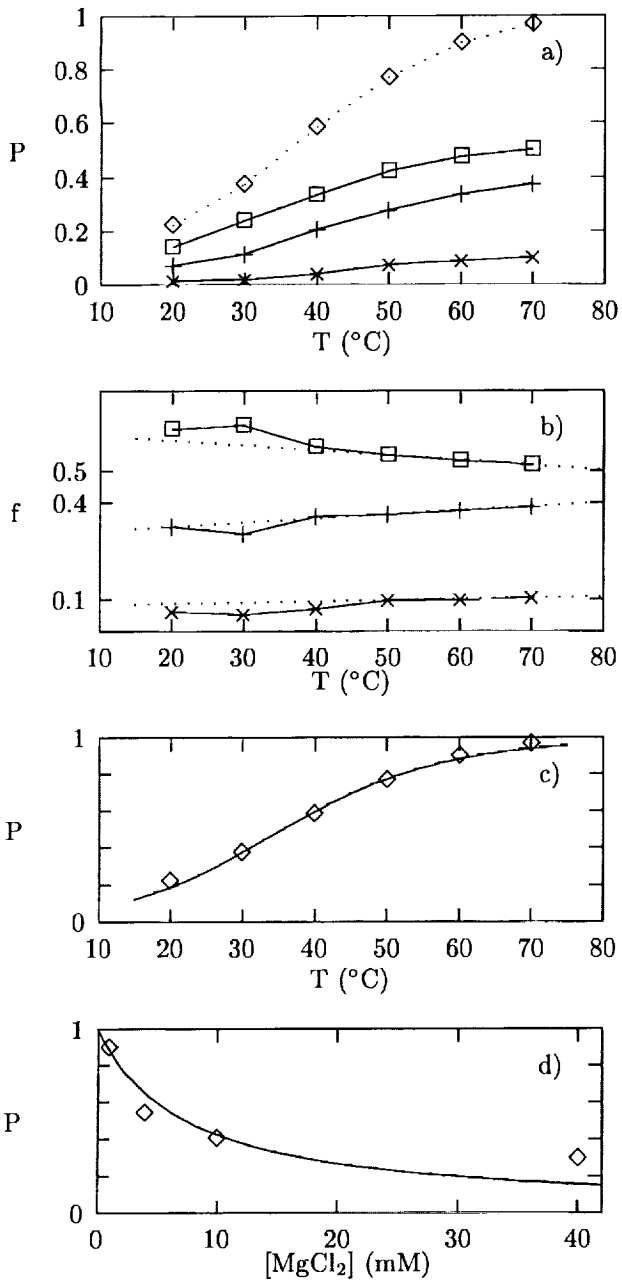

FIG. 1. a) Temperature dependence of termination probability P for phage λ terminator tR2 reported by Wilson and von Hippel.[2] +, □, and × denote the probabilities to terminate 7, 8, and 9 nucleotides downstream from the beginning of the poly-T stretch. The sum is shown as ◇. b) +, □, and × above divided by ◇ to make a branching fraction f. c) Comparison of ionization model Eq. (1) with ◇ from a). The ionization energy has been fit to $\epsilon_0 = 0.7eV$ (16 kcals/mole) and the quantity $n/M^{3/2}$ adjusted to make the curves match at 30 °C. d) Prediction of Eqn. (1) for dependence on Mg^{2+} concentration compared with data of Reynolds et al.[7]

report that the polymerase "will not elongate", i.e., has stopped working properly. This suggests that the effect has something to do with the overall mechanical integrity of the enzyme rather than the termination process alone. Guided by this observation we note that the activated behavior identified by Wilson and von Hippel is actually the formula for conventional monomolecular chemical equilibrium. The probability for a particle of mass M with a binding energy of E_0 to be ionized off the polymerase is

$$P = \frac{1}{1 + Ze^{E_0/k_BT}\, n\, \lambda_{th}^3} \qquad (\lambda_{th} = \sqrt{\frac{2\pi\hbar^2}{Mk_BT}}) \ , \quad (1)$$

where n is the concentration of this component and Z is the change to the internal partition function that results from binding. If one makes the approximation that λ_{th} is a slowly-varying function of temperature and can thus be taken to be constant then this reduces to the formula with which Wilson and von Hippel fit their data.[2] That it works may be seen in Fig. 1c, where we plot the total termination probability from experiment against Eq. (1) with $E_0 = 0.7eV$ and Z adjusted to match experiment at T = 30 °C. Thus reinterpreting this effect as an ionization equilibrium, we may account for the high-temperature intercept and weak temperature dependence seen in Fig. 1b in the following way: In addition to the ionization state the polymerase possesses an internal configurational memory with a number of states of order 10. These code for termination at sites 7, 8 or 9. In the equilibration step, the polymerase molecules come to thermal equilibrium and a fraction P of them become ionized. All of these terminate at one of the three sites when launched. The rest read through.

A candidate for the ionizable component is an Mg^{2+} ion. In their studies of the effects of ion concentrations on termination efficiency, Reynolds et al[7] discovered that Mg^{2+} has the strange and unique effect of increasing termination efficiency to 100% for all terminators studied when reduced below 1 mM. The Mg^{2+} concentration in the experiments shown in Fig. 1d was 10 mM.[2] Extrapolating at T = 30 °C[53] using Eq. (1) we obtain, with no adjustable parameters, the fit to the $[MgCl_2]$ dependence found by Reynolds et al[7] shown in Fig. 1d. The quality of this fit suggests that Mg^{2+} has a special function in regulating transcription, and that the temperature dependence in Fig. 1a is simply a thermal binding relation for this ion. This is corroborated by the recent structural studies of Zhang et al,[54] who report that polymerase crystallized out of 10 mM solution of $MgCl_2$ has a Mg^{2+} ion bound at what appears to be the catalytic site of the enzyme.

There is evidence for more termination channels other than the ionization of Mg^{2+}. In Fig. 2 we reproduce results of Reynolds et al[7] showing that terminator efficiencies tend to saturate at large Mg^{2+} concentration to values other than zero. The saturation values are balanced, and there is an evident tendency of them to cluster. Both effects are consistent with the polymerase executing an instruction at the terminator to read through conditionally, even when the ionizable component is bound, if its memory is appropriately set. There is obviously not

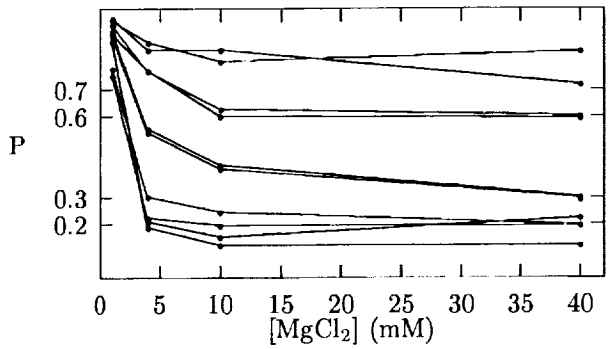

FIG. 2. Termination efficiency as function of $[MgCl_2]$ for 10 terminators, as reported by Reynolds et al.[7] The terminators are, top to bottom at the right edge, *RNA I*, *T7Te*, *rrnB T1*, *trp a L126*, *trp a*, *tR2*, *T3Te,P14*, *tonB t*, and *trp a L153*.

enough data here to draw such a conclusion, however. We note that Reynolds et al[7] also found order-1 effects on the termination efficiency from Cl^- and K^+, although with the opposite sign. The function of these ions is not yet known.

VI. ANTITERMINATION

What experiments can detect internal memory? In general, one would look for cases in which polymerase acts differently under apparently identical conditions, suggesting an internal control mechanism of some kind. Such thinking motivates the following hypothetical experiment: one constructs a template with promoter P followed by two identical terminators and flanking DNA sequences in succession. If termination is stochastic, then the branching ratio at T_2 will be the same as that at T_1. If termination is deterministic and hysteretic, then the branching ratios will be different, depending on details. A passive termination at T_1 would result in no termination at T_2, since the polymerase that reads through has been "polarized", i.e., selected for the memory setting that codes for read-through. An active termination at T_1 would reprogram the memory there and cause a termination probability at T_2 different from that of T_1 but not necessarily zero. Variations of this design, e.g., adding more terminators, combining different terminators, changing their order, etc., could, in principle, answer more sophisticated questions, such as whether and how polymerase is reprogrammed in active read-through and whether non-equilibrium effects are important.

A few such experiments have already been performed on DNA templates containing antiterminators (sequences upstream of terminators that reduce termination efficiencies) and are thus less general than one would like, but they strongly support the idea of polymerase memory. There is indirect evidence in the case of N-antitermination of phage λ, the case most studied, that the memory is a physical attachment of the transcribed mRNA to the polymerase to form a loop.[55] There is also evidence that it is not true generally.[44]

Sequence	T7Te	trp a
AATTGTGAGCGGATAACAATTTCACACAGGAAACAGGGAA	61	99
AATTGTGAGCGGATAACAATTTCACACAGGAAACAGAA..	51	52
AATTGTGAGCGGATAACAATTTCACACAGGAA...	73	99
AATTGTGAGCGGATAACAATTTCACGGAA...	45	99
AATTGTGAGCGGATAACAATTTCAGGAA...	71	99
AATTGTGAGCGGATAACAATTTCGGAA...	75	66
AATTGTGAGCGGATAGGAA...	88	75
No Antiterminator	99	80

TABLE V. Sequences and corresponding termination probabilities at downstream *T7Te* and *trp a* for modified *lac* antiterminators reported by Telesnitsky and Chamberlin.[44]

Sequence	oop t	rpoC t
AAATCTGATAATTTTGCCAATGTTGTACGGAATTC	37	22
AAATCTGATAATTTTGCCAATGTTGGGAATTC...	45	17
AAATCTGATAATTTTGCCAATGTTGGAATTC...	31	19
AAATCTGATAATTTTGCCAATGGAATTC...	29	16
AAATCTGATAATTTTGCCGGAATTC...	25	18
AAATCTGATAATTTGGAATTC...	17	20
AAATCTGATAATTGGAATTC...	15	22
AAATCTGATAATGGAATTC...	11	20
AAATCTGATAAGGAATTC...	19	21
AAATCGGAATTC...	20	16

TABLE VI. Antiterminator sequences constructed by Goliger et al[51] from a promoter from phage 82, together with the readthrough probabilities *in vitro* for downstream terminators *oop t* and *rpoC t*. Note that these terminators are not in series. The underlined sequence on the right is the EcoRI linker.

In 1989 Telesnitsky and Chamberlin[44] reported memory effects associated with the *lac* antiterminator found just downstream of the *Ptac* promoter in *E. coli*. Their key result is reproduced in Table V. Insertion of *lac* 353 nucleotides upstream of the terminator makes different order-1 modifications to the termination efficiences of *T7Te* phage and *trp a*. The antiterminator contains a palindrome, and the antitermination effect is sensitive to modifications of the downstream 15-base-pair sequence. 3 copies of *T7Te* placed in tandem downstream of *lac* showed that the antitermination effect is partially remembered through multiple terminators: the efficiencies were 44%, 60%, and 90%, but without the antiterminator they were 90%, >90%, and >90%.

In another experiment *in vitro* reported in 1989, Goliger et al[51] found that the *E. coli* terminator *rpoC t* and phage terminators *oop t* and t_{82} were strongly antiterminated by a sequence they constructed accidentally. Their key result is reproduced in Table VI. A phage 82 promoter was fused onto a sequence containing either *rpoC t* alone or *oop t*

Sequence	Name
GAGCGCGGCGGGGTTCAGGATGAACGGCAATGCTGCTCATTAGC	putL
GCGTGGTCAAGGATGACTGTCAATGGTGCACGATAAAAACCCA	putR

TABLE VII. Antitermination sequences *putL* and *putR* from the Hong Kong phage HK022.[56]

and *rpoC t* in tandem using the EcoRI linker sequence GGAATTC. This resulted in unexpected antitermination *in vitro* of both terminators, but of different sizes that depended sensitively on the insertion point. The read-through effects in the tandem experiments were unfortunately poorly documented. One can see from Table V that the phage terminator responded more strongly in this experiment than did *rpoC t*. However, the reverse was the case in another experiment in which the antiterminator was a portion of the 6S RNA gene downstream of a phage λ pR′ promoter, and in which factor *NusA* was present. As a control, this latter experiment was rerun with the phage terminator t_{82}, which terminated at greater than 98% in all cases, seemingly immune to antitermination.

King et al[52] reported in 1996 that the *putL* and *putR* antitermination sequences of the Hong Kong phage IIK022,[56] shown in Table VII, caused downstream readthrough of a triple terminator consisting of tR′ from phage λ followed by the strong *E. coli* ribosome operon terminators *rrn B* t_1 and *rrn B* t_2. This effect was sensitive to the choice of promoter. When *putL* was inserted between the *Ptac* promoter and the triple terminator 284 nucleotides downstream and studied *in vivo* the termination probability was 50%. Substituting the phage λ P_L promoter for *Ptac* under the same conditions resulted in complete readthrough (though with wide error bars). When this experiment was repeated *in vitro* the antitermination effect was found to be smaller and to persist through all three terminators. The read-through probabilities at tR′ were 34% and 31% for promotion by P_L and *Ptac*, respectively, but 57% and 27% for *rrnB* t_1 and 76% and 40% for *rrnB* t_2. This result is incompatible with statistical termination, for both the antitermination effect itself and the changes resulting from switching promoters are order-1 effects that do not add. They also reported that reduced Mg^{2+} concentration destroys the antitermination effect.

VII. CONCLUSION

In summary we find that the theory of stochastic termination, which requires natural selection to engineer a physical miracle of balanced rates, is flawed, but that there is ample evidence of a sophisticated and as-yet poorly understood regulatory system in RNA polymerase involving hysteresis, metastability, and long-term configurational memory, all robust phenomena in inanimate matter. On this basis we predict that branching ratios of identical terminators in series will differ by order-1 amounts very generally - specifically in the absence of looping. We propose that the confusion surrounding the existence of polymerase memory is symptomatic of the larger problem that measurement of physical activity on the length and time scales appropriate to life has thus far been impossible, and that overcoming this problem should be one of the high-priority goals of modern nanoscience.

This work was supported primarily by NASA Collaborative Agreement NCC 2-794. SL would like to acknowledge informative discussions with Brian Ring. RBL wishes to express special thanks to the organizers of the Keystone conferences, particularly R. Craig, for encouraging interdisciplinary science. We also wish to thank the Institute for Complex Adaptive Matter at Los Alamos, the Brown-Botstein group at Stanford for numerous stimulating discussions, and H. Frauenfelder, J. W. Roberts, D. Botstein and M. J. Chamberlin for a critical reading of the manuscript.

[1] P. H. von Hippel and T. D. Yager, Science **255**, 809 (1992); ibid., Proc. Natl. Acad. Sci. **88**, 2307 (1991).
[2] K. S. Wilson and P. H. von Hippel, J. Mol. Biol. **244**, 36 (1994).
[3] R. Landick and C. L. Turnbough, Jr. in *Transcriptional Regulation*, ed. by S. L. McKnight and K. R. Yamamoto (Cold Spring Harbor Press, Plainview, N.Y., 1992), p. 407; F. Lee and C. Yanofsky, Proc. Natl. Acad. Sci. **74**, 4365 (1977).
[4] V. Brendel, G. H. Hamm, and E. N. Trifonov, J. Biomol. Struct. Dyn. **3**, 705 (1986).
[5] Y. d'Aubenton, J. Brody, and C. Thermes, J. Mol. Biol. **216**, 835 (1990).
[6] *E. coli* K-12 MG 1655, ftp://ncbi.nlm.nih.gov/genbank /genomes/bacteria/Ecoli/ecoli.ptt (1999); ibid. /ecoli.fna (1999); http://www.ncbi.nlm.nih.gov/BLAST (1999).
[7] R. Reynolds et al, J. Mol. Biol. **224**, 31 (1992); R. Reynolds and M. J. Chamberlin, ibid., 53 (1992).
[8] J. Briat et al, J. Mol. Biol. **198**, 43 (1987).
[9] J. C. McDowell et al, Science **266**, 822 (1994).
[10] S. P. Lynn., L. M. Kasper, and J. F. Gardner, J. Biol. Chem. **263**, 472 (1988).
[11] K. S. Wilson and P. H. von Hippel, Proc. Natl. Acad. Sci. **92**, 8793 (1995).
[12] G. E. Christie, P. J. Farnham, and T. Platt, Proc. Natl. Acad. Sci. **78**, 4180 (1981).
[13] R. F. Fisher and C. Yanofsky, J. Biol. Chem **258**, 8146 (1983).
[14] P. W. Anderson, Science **177**, 393 (1972); R. B. Laughlin *et al.*, Proc. Natl. Acad. Sci. **97**, 32 (2000).
[15] R. Zallen *The Physics of Amorphous Solids* (Wiley, New York, 1983); *Amorphous Solids*, edited by W.A Phillips (Springer, Berlin, 1981); F. H. Stillinger and T. Weber,

Science **225**, 983 (1984).
[16] P. W. Anderson, B. I. Halperin, and C. M. Varma, Phil. Mag. **24**, 1 (1972).
[17] K. Binder and A. P Young, Rev. Mod. Phys. **58**, 801 (1986).
[18] G. P. Singh et al, Z. Phys. B **55**, 23 (1984).
[19] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[20] L. Chen et al, J. Mol. Biol., to be published (1998).
[21] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[22] F. E. Cohen et al, Science **264**, 530 (1994).
[23] K. E. Neet and G. R. Ainslie, Jr., Meth. Enzymol. **64**, 192 (1980).
[24] C. Frieden, J. Biol. Chem. **245**, 5788 (1970).
[25] H. P. Lu, L. Xun, and X. S. Xie, Science **282**, 1877 (1998).
[26] J.-C. Meunier, J. Buc, A. Navarro, and J. Ricard, Eur. J. Biochem **49**, 209 (1974).
[27] J. Ricard, J. Buc, and J.-C. Meunier, Eur. J. Biochem. **80**, 581 (1977).
[28] M. J. Holroyde, M. B. Allen, A. C. Storer, A. S. Warsy, J. M. E. Chesher, I. P. Trayer, A. Cornish-Bowden, and D. G. Walker, Biochem. J. **153**, 363 (1976).
[29] J. P. Shill and K. E. Neet, J. Biol. Chem. **250**, 2259 (1975).
[30] B. A. Peters and K. E. Neet, J. Biol. Chem. **252**, 5345 (1977).
[31] C. K. Surratt, S. C. Milan, and M. J. Chamberlin, Proc. Natl. Acad. Sci. USA **88**, 7983 (1991).
[32] G. Feng, D. N. Lee, D. Wang, C. L. Chan, and R. Landick, J. Biol. Chem. **269**, 22282 (1994).
[33] D. N. Lee, G. Feng, and R. Landick, J. Biol. Chem. **269**, 22295 (1994).
[34] S. Borukhov, A. Polyakov, V. Nikifornov, and A. Goldfarb, Proc. Natl. Acad. Sci. USA **89**, 8899 (1992).
[35] S. Borukhov, V. Sagitov, and A. Goldfarb, Cell **72**, 459 (1993).
[36] M. Orlova, J. Newlands, A. Das, A. Goldfarb, and S. Borukhov, Proc. Natl. Acad. Sci. USA **92**, 4596 (1995).
[37] M. Chamberlin and P. Berg, Proc. Natl. Acad. Sci. USA **48**, 81 (1962).
[38] B. Krummel and M. J. Chamberlin, J. Mol. Biol. **225**, 221 (1992).
[39] D. C. Straney and D. M. Crothers, Cell **43**, 449 (1985).
[40] K. M. Arndt and M. J. Chamberlin, J. Mol. Biol. **213**, 79 (1990).
[41] B. Krummel and M. J. Chamberlin, J. Mol. Biol. **225**, 239 (1992).
[42] R. E. Kingston, W. C. Nierman, and M. J. Chamberlin, J. Biol. Chem. **256**, 2787 (1981).
[43] T. A. Rozovskaya, A. A. Chenchik and R. S. Beabealashvilli, FEBS Letters **137**, 100 (1982).
[44] A. P. W. Telesnitsky and M. J. Chamberlin, J. Mol. Biol. **205**, 315 (1989); *ibid.* Biochemistry **28**, 5210 (1989).
[45] S. Dissinger and M. M. Hanna, J. Biol. Chem. **265**, 7662 (1990).
[46] J. R. Levin and M. J. Chamberlin, J. Mol. Biol. **196**, 61 (1987).
[47] D. Job, J. M. Soulie, C. Job, J. Theor. Biol. **134**, 273 (1988).
[48] D. A. Erie, T. D. Yager, and P. H. von Hippel, Ann. Rev. Biophys. Biomol. Struct. **21**, 379 (1992).
[49] V. Markovtsov, A. Mustaev, and A. Goldfarb, Proc. Natl. Acad. Sci. USA **93**, 3221 (1996).
[50] R. J. Davenport, G. J. L. Wuite, R. Landick, and C. Bustamente, Science **287**, 2497 (2000).
[51] J. A. Goliger et al, J. Mol. Biol. **205** 331 (1989).
[52] R. A. King et al, Cell **87**, 893 (1996).
[53] M. C. Schmidt and M. J. Chamberlin, J. Mol. Biol. **195**, 809 (1987).
[54] G. Zhang et al, Cell **98**, 811 (1999); R. A. Mooney and R. Landick, ibid., 687 (1999); C. L Poglitsch et al, ibid., 791 (1999); J. Fu et al, ibid., 799 (1999).
[55] J. R. Nodwell and J. Greenblatt, Genes Dev. **5**, 2141 (1991).
[56] J. Oberto, R. A. Weisberg, and M. E. Gottesman, J. Mol. Biol. **207**, 675 (1989).
[57] C. Squires et al, Nucl. Acids Res. **9**, 6827 (1981).
[58] H. Sakamoto, N. Kimura, and Y. Shimura, Proc. Natl. Acad. Sci. **80**, 6187 (1983).
[59] N. Nakajima et al, J. Biol. Chem. **257**, 11113 (1982).
[60] B. Hove-Jensen et al, J. Biol. Chem. **261**, 6765 (1986).
[61] D. Bedwell et al, Nucl. Acids. Res. **13**, 3891 (1985).
[62] J. S. Lee et al, Mol. Gen. Genet. **184**, 218 (1981).
[63] P. Regnier, M. Grunberg-Manago, and C. Portier, J. Biol. Chem. **262**, 63 (1982).
[64] J. E. L. Larsen, B. Albrechtsen, and P. Valentin-Hansen, Nucl. Acids Res. **15**, 5125 (1987).
[65] F. Lee et al, J. Mol. Biol. **121**, 193 (1978).
[66] A. M. Wu , G. E. Christie, and T. Platt, Proc. Natl. Acad. Sci. **78**, 2913 (1981).
[67] G. Ralling and T. Linn, J. Bareriol. **158**, 279 (1984).
[68] J. F. Gardner, Proc. Natl. Acad. Sci. **76**, 1706 (1979).
[69] S. R. Wessler and J. M. Calvo, J. Mol. Biol. **149**, 579 (1981).
[70] C. L. Turnbough, K. L. Hicks, and J. P. Donahue, Proc. Natl. Acad. Sci **80**, 368 (1983).
[71] D. Frieden, T. Newman, and M. Freundlich, Proc. Natl. Acad. Aci. **79**, 6156 (1982).
[72] G. Fayat et al, J. Mol Biol. **171**, 239 (1983).
[73] R. P. Lawther and G. W. Hatfield, Proc. Natl. Acad. Sci. **77**, 1862 (1980).
[74] R. A. Young, J. Biol. Chem. **254**, 12725 (1979).
[75] G. L. Duester and W. M. Holmes, Nucl. Acids Res. **8**, 3793 (1980).
[76] rrnft.
[77] T. Grundstroem and B. Jaurin, Proc. Natl. Acad. Sci. **79**, 1111 (1982); B. Jaurin et al, Nature **290**, 221 (1981).
[78] C. M. Joyce and N. D. F. Grindley, J. Bact. **152**, 1211 (1982).
[79] K. Postle and R. F. Goode, Proc. Natl. Acad. Sci. **80**, 5235 (1983).
[80] A. Y. Cheung and D. Söll, J. Biol. Chem. **259**, 9953 (1984).
[81] 120
[82] B. Jaurin et al, Nature **290**, 221 (1981).
[83] G. Zurawski et al, Proc. Natl. Acad. Sci. **75**, 4271 (1978).
[84] P. Verde et al, Nucl. Acids. Res. **9**, 2075 (1981).
[85] J. Brosius et al, J. Mol Biol. **148**, 107 (1981).
[86] J. Brosius, Gene **27**, 161 (1984).
[87] H. Horowitz and T. Platt, J. Biol. Chem. **257**, 11740 (1982).

[88] A. M. Easton and S. R. Kusher, Nucl. Acids. Res. **11**, 8625 (1983).