

# Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase

CRAIG TUERK AND LARRY GOLD

**High-affinity nucleic acid ligands for a protein were isolated by a procedure that depends on alternate cycles of ligand selection from pools of variant sequences and amplification of the bound species. Multiple rounds exponentially enrich the population for the highest affinity species that can be clonally isolated and characterized. In particular one eight-base region of an RNA that interacts with the T4 DNA polymerase was chosen and randomized. Two different sequences were selected by this procedure from the calculated pool of 65,536 species. One is the wild-type sequence found in the bacteriophage mRNA; one is varied from wild type at four positions. The binding constants of these two RNA's to T4 DNA polymerase are equivalent. These protocols with minimal modification can yield high-affinity ligands for any protein that binds nucleic acids as part of its function; high-affinity ligands could conceivably be developed for any target molecule.**

**U**NDERSTANDING SITE RECOGNITION BY A NUCLEIC ACID binding protein has historically involved the collection and comparison of a number of binding sites from which a consensus sequence can be derived. The existence of many naturally occurring binding sites for a specific protein assembly, such as in transcriptional initiation, simplifies the analysis (1, 2). When there is only one binding site per genome for a protein, such as for the nucleation target of the coat protein of the RNA bacteriophage R17, the analysis of what nucleotides are required for binding can be addressed by creating and testing a number of individual mutants (3). Although studies of such RNA-protein interactions have been simplified by the development of oligonucleotide-directed *in vitro* transcription with T7 RNA polymerase (4), many nucleic acid targets are too extensive and complex to be so analyzed. In addition, even in the relatively simple binding site of R17 coat protein, the number of individual sequences necessary to do a thorough study of all possible sequences is prohibitively large.

We have been studying the interaction between the bacteriophage T4 DNA polymerase (gp43) and the ribosome binding site of the mRNA that encodes it. The binding of gp43 at a site that overlaps

the Shine and Dalgarno sequences prevents translational initiation and achieves autogenous regulation (5). The minimally sized target is a 36-nucleotide (nt) fragment, which includes a 5-bp helix with 8 loop nucleotides (Fig. 1); the distantly related bacteriophage RB69 also has an RNA hairpin at the ribosome binding site of its DNA polymerase gene. The loop sequence of the RB69 hairpin is identical to that of T4 (6). In order to elucidate what is responsible for the loop nucleotide bias in this translational operator interaction, we have developed a novel method for rapidly selecting preferred binding sequences from a population of random sequences. We call this method systematic evolution of ligands by exponential enrichment (SELEX).

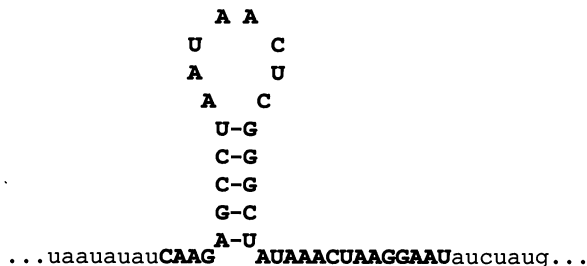
**The SELEX procedure.** The method relies on mechanisms usually ascribed to the process of evolution, that is, variation, selection, and replication. A pool of RNA's that are completely randomized at specific positions is subjected to selection for binding, in this case to gp43 on nitrocellulose filters. The selected RNA's are amplified as double-stranded DNA that is competent for subsequent *in vitro* transcription. This newly transcribed RNA is enriched for better binding sequences and is then subjected to selection to begin the next cycle. Multiple rounds of enrichment result in the exponential increase of the best binding ligands until they dominate the population of sequences.

The first step in these experiments was to create a pool of variant sequences from which RNA ligands of relatively high affinity for gp43 could be selected through multiple rounds (Fig. 2). A 110-base single-stranded DNA template for *in vitro* transcription was created by ligation of three synthetic oligonucleotides in the presence of two bridging oligonucleotides (Fig. 2, step a): One of the template-creating oligonucleotides (oligo 5) is also used as the 3' primer in (i) reverse transcription of the *in vitro* transcripts, (ii) subsequent amplification by the polymerase chain reaction (PCR) (7), and (iii) sequencing of transcripts and plasmid DNA. One of the bridging oligonucleotides (oligo 1) contains the information required for T7 RNA polymerase transcriptional initiation and sufficient sequence complementarity to the complementary DNA (cDNA) of the *in vitro* transcript to serve as the 5' primer in the PCR amplification steps. The ligated template encodes a 92-nt transcript that contains the entire RNA recognition site for T4 DNA polymerase (36 nt) flanked by primer annealing sequence information for the 5' and 3' oligonucleotides used in PCR (8). In place of the sequence that would encode the wild-type hairpin loop sequence AAUAACUC (in oligo 4 of Fig. 2, step a), we incorporated completely random sequences at these eight positions so that there are theoretically 65,536 individual species (9). We also synthesized a template carrying the wild-type loop sequence and compared it with the sequence of the starting variable RNA (Fig. 3).

The authors are in the Department of Molecular, Cellular, Developmental Biology, University of Colorado, Boulder, CO 80309.

The dissociation constant ( $K_d$ ) for the complex of gp43 with the wild-type sequence in this context is about  $5 \times 10^{-9}$  M, while for the population of variable sequences the dissociation constant is about  $3.2 \times 10^{-7}$  M. Thus the average binding of all of the species in the population of variable sequences is about 60 times less than that of wild type, but about 100 times better than nonspecific binding (6). Our goal was to enrich the population for the highest affinity ligands.

In vitro transcripts of variable loop sequence were mixed with purified gp43 at three different ratios of RNA to protein throughout the multiple rounds of selection on nitrocellulose filters (Fig. 2, step c). The concentrations of gp43 for experiments A, B, and C



**Fig. 1.** The gene 43 translational operator in bacteriophage T4. Shown is the sequence at the ribosome binding site of gene 43. The bold-faced capitalized letters indicate the extent of the information required for binding of gp43 (6).

**Fig. 2.** Systematic evolution of ligands by exponential enrichment. (step a) The five oligonucleotides, numbered 1 through 5, were synthesized as described (39), and have the following sequences:

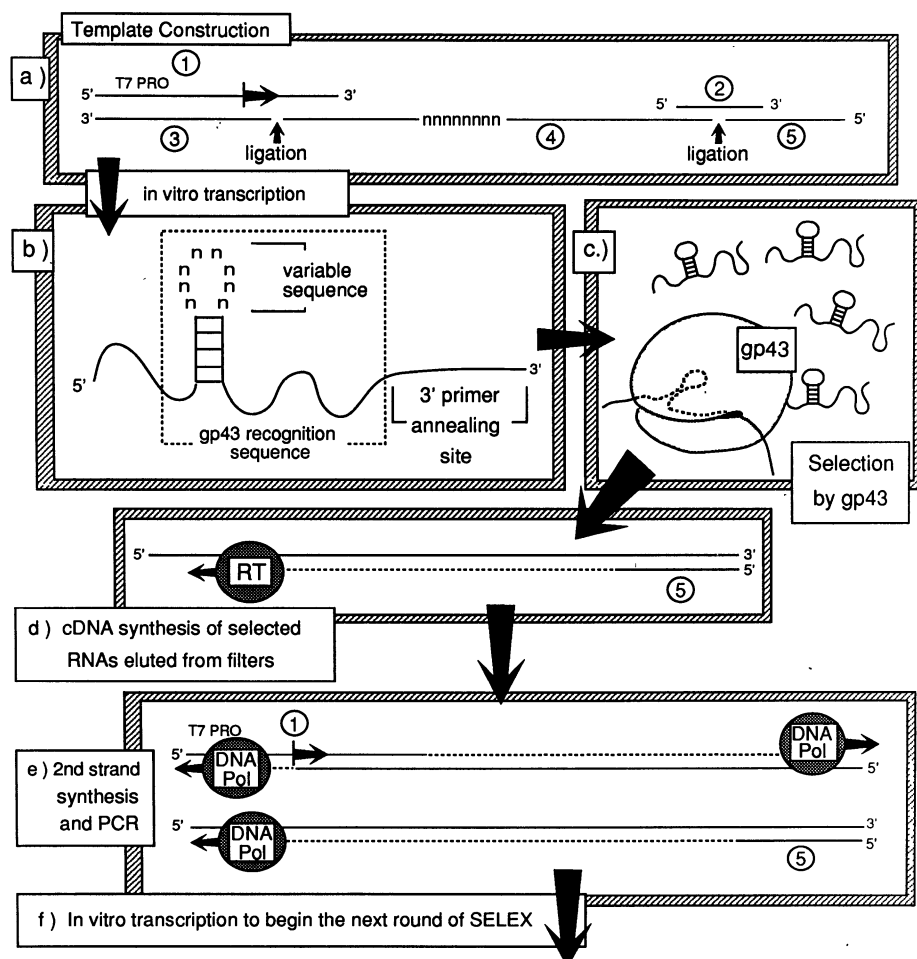
- 1) 5'-TAATACGACTCACTATAGGAGGCCAACACCAATTCGAATCAAG-3'
- 2) 5'-GGGCTATAAACTAAGGAATATCTATGAAG-3'
- 3) 5'-GAATTGGTGTGGTCCCTATAGTGAGTCGATTA-3'
- 4) 5'-ATATCTCTTAGTTTATAGCCCNNNNNNNAGGCTCTTGATTG-3'
- 5) 5'-GTTTCAATAGAGATATAAAATTCCTTCATAG-3'

The oligonucleotides hybridize as shown and were ligated according to Sambrook *et al.* (11). (b) Approximately 500 pmol of the resultant 110-base template was gel purified, annealed to equal amounts of oligo 1, and was used in transcription reactions in vitro as described (4) from which 7.5 nmol of gel-purified transcript was recovered. (c) The resultant transcripts were gel purified and subjected to selection on nitrocellulose filters by gp43. The RNA and protein were mixed in the concentrations shown in Table 1 in a 100- $\mu$ l volume, incubated at 37°C for 4 minutes, and washed through a nitrocellulose filter; the RNA was eluted from the filters (6). The amount of RNA recovered in the absence of protein was about 0.1 percent of the input RNA. The average efficiency of recovery of RNA from the filters is approximately 80 percent. (d) Complementary DNA copies of the selected RNA's were made by AMV reverse transcriptase synthesis primed by oligo 5 (50 pmol in 50- $\mu$ l reactions) as described (39). (e) The cDNA's were amplified by *Taq* DNA polymerase chain extension of oligo 1 (which carries the essential T7 promoter information as well as complementing the cDNA) and oligo 5. To the cDNA synthesis reaction was added 50 pmol of oligo 1 and in a reaction volume of 100  $\mu$ l was amplified with *Taq* DNA polymerase as described by Innis *et al.* (7) for 30 cycles. (f) The double-stranded DNA products of the amplification reaction were transcribed in vitro and used in the next round of selection (step c). In vitro transcription was performed on these amplified templates in 50- $\mu$ l volumes as described (4). Deoxyribonuclease I

were 0.03, 0.03, and 0.3  $\mu$ M, respectively. The concentrations of RNA for experiments A, B, and C were 0.3, 30, and 30  $\mu$ M, respectively, making the ratios of RNA to gp43 10, 1000, and 100, respectively. During all three experiments the RNA was in excess of a fixed number of binding sites (gp43) so that there was substantial competition between potential ligands during the selection process. Typically the amounts of RNA retrieved from the filters were roughly equivalent to the amount of gp43 in the binding reaction, as would be expected from these conditions of RNA saturation.

The RNA harvested from filters was used as template for cDNA synthesis and subsequent amplification with *Taq* DNA polymerase (Fig. 2, steps d and e). Because the 5' primer in this amplification contained the information specifying initiation by T7 RNA polymerase, the resultant DNA could be used as templates to prepare in vitro transcripts for the next round of selection on nitrocellulose filters by gp43. Less than one-tenth of the RNA from this step was used in the next cycle; the remainder was used to trace the history of the selection.

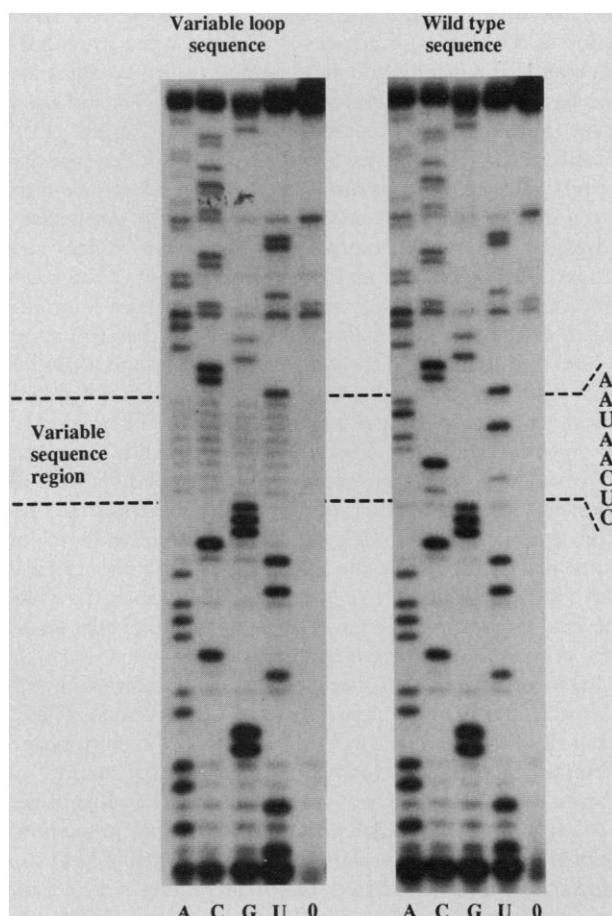
**The selection of a batch "consensus" sequence.** We monitored the progress of the selection by filter binding assays of labeled in vitro transcripts derived from each PCR reaction (6). After the fourth round of selection and amplification, the labeled products exhibited binding to gp43 which was indistinguishable from that of the wild-type RNA. The unused RNA products from each round for one experiment (Fig. 4B) and from the fourth round (of selection, amplification, and transcription) for all three experiments (Fig. 5) were gel-purified and sequenced. The sequence of the



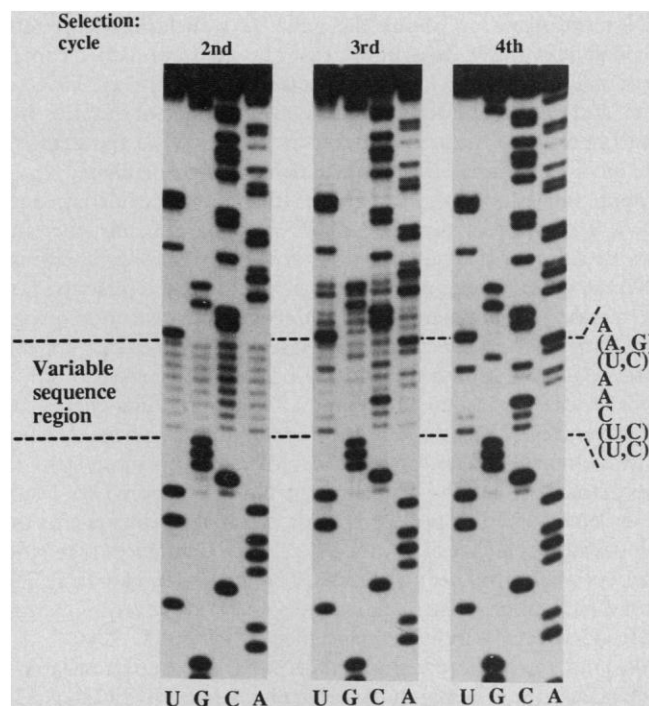
(DNase I) was added (5 units at 37°C for 15 minutes; Pharmacia) to remove the DNA template and the mixture subsequently was extracted with phenol and precipitated with ethanol as described (11).

purified in vitro transcripts derived from the second, third, and fourth rounds of selection and amplification for experiment B showed no bias until after the third selection (Fig. 4). At this point there was a detectable bias that was complete by the end of the fourth round for the apparent consensus sequence A(A,G)(U,C)AAC(U,C)(U,C). Two further rounds failed to produce any noticeable change in the pattern of these batch sequences. Sequencing of the RNA transcribed after the fourth selection and amplification for A, B, and C (Fig. 5), independently, gave a similar consensus (10). The wild-type sequence, AAUAACUC, is contained in this consensus. By comparing normalized band intensities on sequencing gels with the Ambis™ radioanalytic imaging system, we determined that the bias for wild type at the varied positions in this consensus is on average 1.4 times that of the alternative for experiment A, 1.1 for experiment B, and 1.6 for C.

**Two predominant sequences make up the “consensus” sequence.** As a first approximation, the sequencing of the selected batches of RNA species provided a consensus sequence reminiscent of those found for other nucleic acid binding moieties; that is, there were certain nucleotides that were rigidly specified, and certain positions in which either purines or pyrimidines are exclusively required. Thus the consensus, ARYAACYY, where R is A or G, and Y is U or C, could conceivably contain 16 different sequences [or more, if minor variants existed in the population (see below)]. We realized that some of these 16 would hypothetically form a different



**Fig. 3.** Sequence comparison between the batch variable loop and wild type. The transcripts containing the variable loop batch of sequences and wild-type loop sequence were gel-purified, and the products were subjected to electrophoresis and dideoxy sequencing as described (39). The RNA complement to the specific dideoxynucleotide used in each reaction is indicated below each lane; in the lane marked 0 extension reactions were done in the absence of dideoxynucleotides.



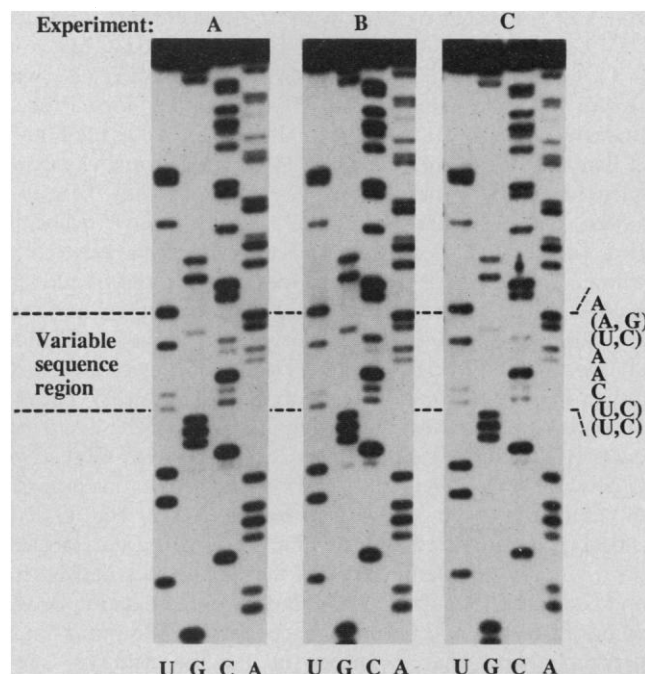
**Fig. 4.** Batch RNA sequences through rounds of selection. Sequencing and labeling were done as in Fig. 3. The transcripts were derived from amplified products of selections 2, 3, and 4 for experiment B.

loop conformation, that is, extra base pairs, than we had predicted to exist in the wild-type sequence. To find out what allowable combinations of nucleotides actually existed within the apparent consensus sequence, we cloned individual sequences from the fourth round batch of experiment B. We chose experiment B because there appeared to be an even distribution of the two allowable nucleotides at each of the four variable positions, which was confirmed by the scanning data, and we wanted to see the greatest variability in sequences that were selected for binding. We amplified cDNA's of this batch by PCR with two cloning oligonucleotides (5'-CCGAAGCTTCTAATACGACTCACTATAGGGAG-3' and 5'-CCGGATCCGTTTCAATAGAGATATAAAATTC-3') which provided 5' Hind III and 3' Bam HI restriction endonuclease sites to the amplified sequences. Individuals were cloned at the Hind III and Bam HI sites of pUC18 (11). Twenty individual clones were sequenced. None of these clones were mutant at any place in the operator sequence other than at the positions that we deliberately varied. Surprisingly, the batch sequence is actually composed of two predominant sequences (Fig. 6). One is the wild-type sequence, AAUAACUC, of which nine were isolated. The other, AGCAACCU (which we call the major variant), is mutant at four positions and was present in eight clones. The ratio between these two numbers fit the ratio determined by the Ambis data well. The other three sequences were single transition mutations of these two major sequences (AAUGACUC, AAUAACUU, AGCGACCU). [Notably, none of the isolated clones contain a C at the fourth position (10), although we did retrieve two mutant species with G at this position.] We conducted filter binding experiments with labeled in vitro transcripts containing each of these sequences derived from Bam HI-digested templates. We found a close correlation between binding affinity and selected abundance (Fig. 7). Additional binding experiments show that the major variant and wild-type species compete against each other for binding to gp43 and the dissociation rate constant for wild-type species is about twice that of the major variant (Fig. 8).

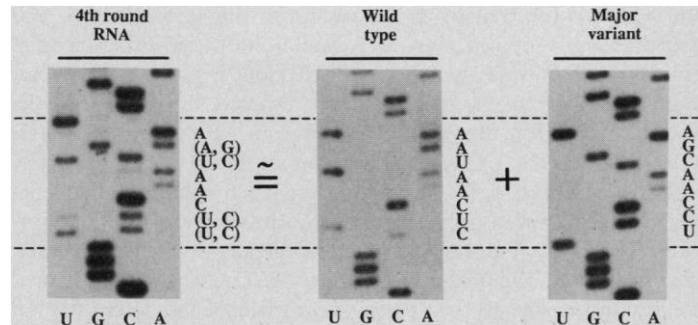
### New information about the gene 43 translational operator.

We had previously determined that the operator RNA contains three major features: a loop of eight nucleotides, a stem of five base pairs, and single-stranded sequences on either side of the stem (four nucleotides 5' to the stem and 14 nucleotides 3' to the stem) (6). The binding affinity of the helix and flanking single-stranded regions is approximately 100 times greater than nonspecific binding to RNA. The essential stem and single-stranded domains, therefore, provide substantial binding to the protein in the loop-randomized RNA, as though these two fixed domains directly interact with the polymerase and position the variable sequences at the loop recognition domain. What was surprising, at least initially, is that enhancement in binding provided by a loop of eight nucleotides is only 60 times greater (or just an additional 2.5 kcal in binding energy) and yet is completely conserved in the distantly related phage RB69. In contrast, many differences are allowed in the operator stem and single-stranded flanking sequences in RB69 compared to T4, yet these elements of the operator provide 3 kcal of binding energy over that of nonspecific binding in T4 (6). (Such differences may reflect interaction with parts of the RB69 polymerase that vary from T4's, however.) Thus we have a paradox, namely conservation of many nucleotides for a marginal enhancement in binding.

We undertook the evolution experiment to examine what loop nucleotides were important for sequence-specific binding. The surprising answer is that two different loop sequences yield operator RNA's with nearly identical binding constants, and that few nucleotide substitutions are tolerated in either loop. Furthermore, conventional wisdom about the two allowed loops would suggest that the major variant species has an extended stem (by two base pairs) and a smaller loop (four rather than eight residues). Clearly structural work on the two RNA species must be done, both in solution and as complexes with gp43. The powerful result can be confronted immediately: two sequences (from a calculated pool of 65,536 sequences) provide maximal binding to gp43, and yet provide only a marginal enhancement of binding. Even more surprising, given conventional wisdom about the secondary structure of RNA, is that



**Fig. 5.** Batch RNA sequences derived from the fourth round of selection for experiments A, B, and C (see text for concentrations of protein and RNA). Labeling and conditions are as in Fig. 3.



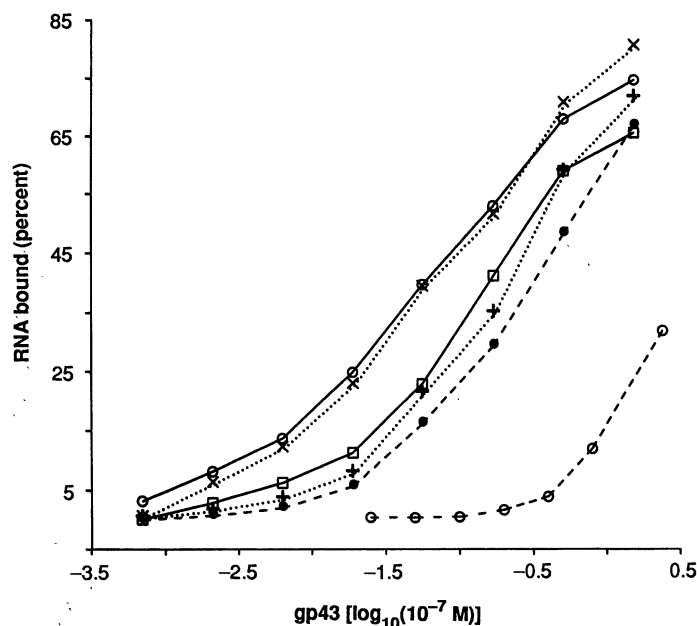
**Fig. 6.** The clonal composition of the "consensus" sequence. Of the three sequencing gels, the first is the batch sequencing of selected RNA's as shown in Fig. 4 for experiment B; the second and third gels show sequencing products derived from plasmid templates (26) of representative clonal isolates of the batch population. These two individual sequences alone yield the batch "consensus" that is observed.

the major variant species is predicted to have two extra base pairs whose composition in primary sequence is specifically required: it is likely that other base pair combinations at the same positions were tested during the evolution experiment yet no other configurations emerged.

### Models of equivalence for wild-type and major variant RNA.

The equivalence of yield by the SELEX procedure and binding affinity for the wild-type and major variant sequences raises the question of how this equivalence is translated at the structural level. It is possible that the bound wild-type sequence contains an A-C pair as has been demonstrated in DNA duplexes (12) and argued to exist by phylogenetic comparison of various tRNA's (13). The observed specification of the base-paired nucleotides may involve protein-RNA helix contacts similar to those found in the interaction between glutamyl tRNA and glutamyl tRNA synthetase (14). Alternatively, the extra base pairs of the major variant may be denatured when bound by gp43, as has been observed in cocrystals of the Klenow fragment of *Escherichia coli* polymerase I and primer-template junctions (15). A third alternative is that the respective structures shown in Fig. 8 are indicative of the bound forms of each species, and that each makes equivalent, although not necessarily identical, contacts with the protein. An extreme form of such mimicry by nonidentical structures is found in the 3' termini of many plant virus genomes that are aminoacylated by host tRNA synthetases (16). Finally, each of these loop sequences may participate in unusual base pair combinations with other parts of the operator that correctly fold the RNA, as has been implicated in the tertiary interactions of tRNA's (17). The resolution of these models, which may have some overlap in reality, awaits further structural studies of free and complexed ligands.

**In vitro compared to natural evolution.** Evolution in vitro of RNA species has been extensively studied in one system. The RNA-directed RNA polymerase (or RNA replicase) of bacteriophage QB has been used to perform "Darwinian selection experiments." Seven key papers by Spiegelman *et al.* were aimed at Darwinian selection (18-24). In a test tube RNA replication, dependent on the synthetic capacity of the purified replicase of QB, yielded small RNA's capable of rapid and accurate replication; minor alterations in the conditions of the serial transfers yielded different "winning" RNA species. Kramer *et al.* (24) were able to start with a small RNA of 218 nucleotides and select, from that species, mutants that outperformed the "wild type" when the selective conditions were altered to include 15  $\mu$ M ethidium bromide. Spiegelman's experiments depended on the binding and catalytic activities of the purified replicase; the selected RNA had to be bound efficiently by the replicase to initiate



**Fig. 7.** Nitrocellulose filter binding of labeled RNA's with varying concentrations of gp43. The RNA's were transcribed from templates amplified from individual plasmid clones or in the case of the NNNNNNNN sequence as in step c of Fig. 2. The percentage of input RNA that was retained on nitrocellulose filters was calculated as an average of three measurements taken at each concentration of gp43. The average standard deviation of all of the measurements for each of the RNA species was less than 2 percent of input RNA. The loop sequences and the symbols representing them in the graph are as follows: AAUAACUC (wild type)  $\circ$ — $\circ$ , AGCAACCU (major variant)  $\times$ — $\times$ , AAUAACUU  $\square$ — $\square$ , AGCGACCU  $+$ — $+$ , AAUGACUC  $\bullet$ — $\bullet$ , and NNNNNNNN  $\circ$ — $\circ$ .

replication (of either plus or minus strands) at one end of the evolving template and had to serve as a kinetically favored template during elongation of RNA. That is, the selected RNA's were required to yield fast kinetics for both binding and elongation.

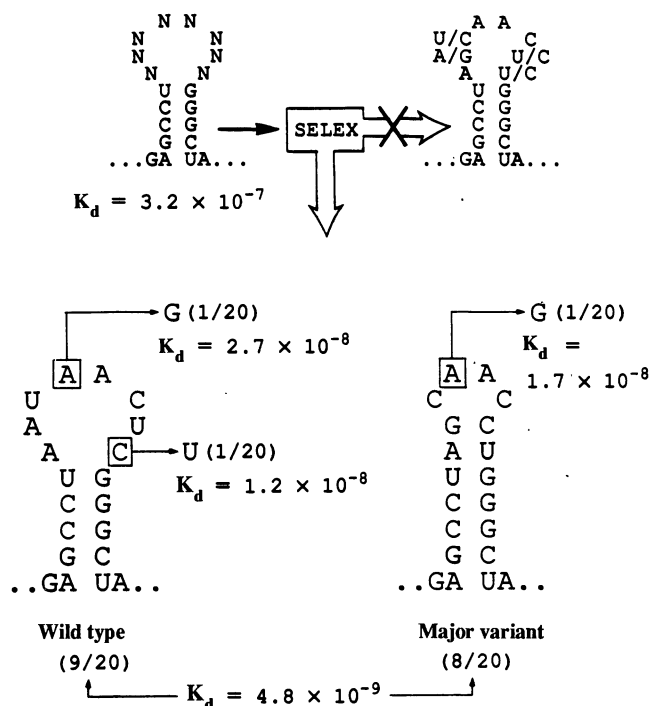
In SELEX, we have separated the binding (initiation) reaction and the elongation reaction. We use binding itself as the selective pressure, and substitute for the elongation reactions of QB RNA replicase the combined reactions of reverse transcriptase, *Taq* DNA polymerase, and T7 RNA polymerase. We thus generalize the Spiegelman paradigm to any molecule that binds nucleic acid. Furthermore, we utilize random nucleic acids for the starting material in the selection, and thus (in the present example) do not depend on the high intrinsic error rate during elongation by QB replicase to sample the widest variety of possible sequences or structures during SELEX. We have designed experiments in which a vast number of possible RNA ligands are sampled at the same time rather than only those ligands that are sequentially available from some arbitrary (or previously selected) initial sequence, which is what appears to happen in natural evolution.

We have shown that DNA polymerase autoregulation is not essential to the growth of T4 and provides no discernible advantage under laboratory conditions (6). However, the operator-repressor interaction does exist and, over the time scale of T4 evolution, must confer some selective advantage. During this natural evolution, sequence variation in different parts of the gene 43 mRNA in and around the translational operator must have been subject to functional constraints of more primary importance. The Shine and Dalgarno sequence is a determinant of translational initiation, yet it is also within the minimal piece of RNA to which gp43 will bind. The sequences upstream from the operator contain signals for transcriptional termination and initiation. The sequences immediately downstream from the operator encode gp43, the DNA

polymerase. The operator sequence that arose, therefore, must have been one whose activity was not compromised by or did not compromise these sequences with more fundamental constraints. In addition, T4 sequences are A-T rich, so that there has been a global bias for A-U rich operator sequences.

The use of SELEX gives us the opportunity to bypass these constraints, although during experimental design we inevitably introduced some bias on what loop sequences would evolve during our experiments (8). In spite of all this, we came up with the same answer that arose through natural evolution. We also got answers that have not been found, and may never be found, in nature. Within the present experimental design, wild type proved to be equivalent to the highest affinity ligand possible. As we continue our analysis of the gene 43 translational operator with SELEX, we have the opportunity to rapidly distinguish between those biases accumulated over the course of natural evolution and those that are important only for binding to gp43. We may select RNA ligands that do not resemble wild type and possibly some that bind better.

**The roots of SELEX.** The philosophy of studying a pool of sequences completely randomized at various positions was developed in our laboratory (25) in order to understand what relation the information content of target sequences had to binding energy in a nucleic acid-protein interaction (26). By cloning with oligonucleotides that were randomized at some nucleotide positions and screening individual transformants, sequence-specific contributions to translational initiation in *E. coli* were studied (25). Subsequently, Schneider and Stormo extended this approach in the study of transcriptional initiation by T7 RNA polymerase (27). The cloning, screening, and sequencing were laborious and time-consuming tasks, however. In addition, the screens often depend on phenotypes that sum a number of intermediate processes and possible in vivo variables in these and other such studies (28–30). SELEX bypasses these caveats, although cloning is still essential to interpretation.



**Fig. 8.** Summary of results. The use of SELEX did not yield the apparent consensus one would expect from the batch sequences shown in Fig. 5, but yielded the wild-type and major variant species with three single mutants. The frequencies of each species in the 20 isolates tested are shown with the approximate  $K_d$ 's derived from the filter binding assays shown in Fig. 7.



Struhl *et al.* have also made extensive use of "random targeted mutagenesis" (28), and similarly bypassed such problems by isolating a consensus collection of double-stranded DNA sequences by affinity chromatography to the yeast transcriptional activator GCN4 (31). Joyce and his colleagues have developed an RNA evolution system that selects for splicing activity (32, 33).

SELEX combines the power of genetic selection techniques with the precision and controllability of biochemical experiments performed *in vitro*. The number of individual molecular species that can be simultaneously tested is much larger than can be directly assessed *in vivo* and we require no scorable phenotype other than binding to the partitioning agent (in this case, gp43). The power of this technique is underscored by its ability to distinguish between species whose binding to gp43 varies by as little as twofold. The theoretical basis of these experiments with consideration of thermodynamic and kinetic parameters and strategies for optimization of these methods is in hand (34).

**Applications.** Among many general applications of this technique, SELEX analyses can be used to determine the optimal binding sequences for any nucleic acid binding protein. The PCR technique can be modified to produce single-stranded DNA, double-stranded DNA, or as we have done, templates for *in vitro* transcription. There also exist isothermal methods of amplification where RNA is both input and output (35) and methods of efficiently mutagenizing during amplification (36) so that continuous variation can be added. All that is further required for SELEX analysis is a relatively pure partitioning agent, often a protein, and some method for separating those nucleic acids that are bound to the agent from those that are not. (Filter binding was adequate in our example, but column or gel retardation may better serve other applications.) SELEX should simplify and accelerate the study of the interactions between transcriptional activators and repressors and transcription complexes at promoter sites, replication accessory proteins and DNA polymerases at origins of replication, and ribosomes and translational repressors at ribosome binding sites.

By extension, any partitioning agent may be used to educate RNA molecules. Small molecules (such as amino acids, nucleotide cofactors, or transition-state analogues) can be bound to insoluble supports to partition RNA's that interact specifically with these substrates. The development of RNA recognition of such "epitopes" would inject new insights into how RNA molecules could have served a wide variety of enzymatic tasks in primordial life-forms, as has been hypothesized (37, 38), and would be the first step in creating (or recreating) these novel (or extinct) RNA enzymes.

We have previously shown that the RNA target of T4 DNA polymerase selectively inhibits its replicative function (6). Thus, the products of SELEX can affect the activity of the protein to which they have been fit. We expect that, at the very least, nucleic acid ligands that inhibit replicative proteins of epidemiologically important infections can be likewise evolved. Other, more sophisticated effects of evolved RNA molecules on target activities may be possible.

Finally, SELEX may be just the beginning of evolution in a test tube. Immunoprecipitation of polysomes to enrich for mRNA's that encode antigens is already a proven technique (11). We expect that, with appropriate modifications, nascent peptides and proteins resident on translating ribosomes can be selected by appropriate targets, and that such selection would partition the ribosome-associated mRNA's from a pool of variants. Mutagenic amplification would create a fresh pool of variant mRNA's that would give rise to selectable different translation products. Over a number of rounds, the proteins would be engineered to best recognize the target of interest. SELEX, therefore, heralds a new era in novel molecular design unrestrained by the rules that govern organismic survival and

replication. SELEX could thus provide unpredictable and unimaginable molecular configurations of nucleic acids and proteins with any number of targeted functions.

## REFERENCES AND NOTES

1. D. Pribnow, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 784 (1975).
2. D. K. Hawley, W. R. McClure, *Nucleic Acids Res.* **11**, 2237 (1983).
3. O. C. Uhlenbeck, J. Carey, P. J. Romaniuk, P. T. Lowary, D. Beckett, *J. Biomol. Struct. Dynamics* **1**, 539 (1983).
4. J. F. Milligan *et al.*, *Nucleic Acids Res.* **15**, 8783 (1987).
5. M. Andrade *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7942 (1988).
6. C. Tuerk, S. Eddy, D. Parma, L. Gold, *J. Mol. Biol.* **213**, 749 (1990).
7. M. A. Innis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9436 (1988).
8. In order to limit selection against loop sequences that base pair with other parts of the molecule, we made the 3' end of the molecule the wild-type T4 sequence, even at the 3' primer annealing site which is dispensable for binding by gp43 (6). The T4 genome is A-T rich, however, and the wild-type sequences that are 5' to the operator region on the gene 43 mRNA are especially so. We have made 5' deletions in this region that replaced wild-type sequence with polylinker sequences that did not reduce the affinity of the encoded RNA's for gp43 (6). In order to ensure the success of the PCR reaction (which is dependent on annealing of the oligonucleotide primers at 72°C) and to limit the possibilities of selecting against sequences merely because they would make competing structures, we designed the part of the 5' oligo that encodes the 5' of the *in vitro* transcript to be C-A rich.
9. In the 500 pmol of original template there would be approximately  $5 \times 10^9$  copies of each original (single-copy) sequence if the incorporation of variable nucleotides in the oligonucleotide synthesis were unbiased. If a particular nucleotide was incorporated at only 60 percent of expected frequency either during random oligonucleotide synthesis or *in vitro* transcription, then the frequency of the homopolymeric eight nucleotide sequence (for example, GGGGGGGG) would be 1 percent of the others. In the lowest concentration of RNA used (0.3  $\mu$ M) there would be approximately three million copies of this hypothetically deficient sequence.
10. There is a stop in lane C at the fourth loop position that is generally weak relative to the A at this position but is noticeably stronger in experiment A. We have not determined whether this is indicative of an allowed C at this position for binding to gp43 or whether this stop is indicative of a marked tendency for misincorporation (especially of dideoxyguanosine) at this position which we have seen in the sequencing of some preparations of wild-type RNA. We did not isolate any cloned sequences that contain a C at this position.
11. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989), sections 1.13, 1.85 to 1.86; *ibid.*, sections 8.9 and 8.10; *ibid.*, sections 1.63 to 1.67; *ibid.*, sections E.10 to E.15, and E.3 to E.4.
12. W. N. Hunter, T. Brown, N. N. Anand, O. Kennard, *Nature* **320**, 552 (1986).
13. J. Ninio, *J. Mol. Biol.* **56**, 63 (1971).
14. M. A. Rould, J. J. Perona, D. Söll, T. A. Steitz, *Science* **246**, 1135 (1989).
15. P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8924 (1988).
16. T. C. Hall, *Int. Rev. Cytol.* **60**, 1 (1979).
17. J. R. Sampson *et al.*, *Biochemistry* **29**, 2523 (1990).
18. D. R. Mills *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 217 (1967).
19. R. Levisohn and S. Spiegelman, *ibid.* **60**, 866 (1968).
20. ———, *ibid.* **63**, 805 (1969).
21. R. Saffhill *et al.*, *J. Mol. Biol.* **51**, 531 (1970).
22. D. L. Kacian *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3038 (1972).
23. D. R. Mills, F. R. Kramer, S. Spiegelman, *Science* **180**, 916 (1973).
24. F. R. Kramer *et al.*, *J. Mol. Biol.* **89**, 719 (1974).
25. J. Childs *et al.*, *UCLA Symp. Mol. Biol.* **30**, 341 (1985).
26. T. D. Schneider *et al.*, *J. Mol. Biol.* **188**, 415 (1986).
27. T. Schneider and G. D. Stormo, *Nucleic Acids Res.* **17**, 659 (1989).
28. A. R. Oliphant and K. Struhl, *Methods Enzymol.* **155**, 568 (1987).
29. ———, *Nucleic Acids Res.* **16**, 7673 (1988).
30. K. T. Min, M. H. Kim, D.-S. Lee, *ibid.*, p. 5075.
31. A. R. Oliphant, C. J. Brandl, K. Struhl, *Mol. Cell. Biol.* **9**, 2944 (1989).
32. G. F. Joyce, in *RNA: Catalysis, Splicing, and Evolution*, M. Belfort and D. Shub, Eds. (Academic Press, New York, 1982), pp. 1–17.
33. D. L. Robertson, and G. F. Joyce, *Nature* **344**, 467 (1990).
34. D. Irvine *et al.*, in preparation.
35. J. C. Guatelli *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1874 (1990).
36. D. Leung, E. Chen, D. V. Goeddel, *Technique* **1**, 11 (1989).
37. H. B. White, in *The Pyridine Nucleotide Coenzymes*, J. Everse, B. Anderson, K. You, Eds. (Academic Press, New York, 1982), pp. 1–17.
38. S. A. Benner, A. D. Ellington, A. Tauer, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7054 (1989).
39. P. Gauss, M. Gayle, R. B. Winter, L. Gold, *Mol. Gen. Genet.* **206**, 24 (1987).
40. We thank G. Stormo and T. Schneider for their very direct and synergistic influence on our thinking during the early years that led up to this work; O. Uhlenbeck and the many people associated with his laboratory, past and present, who developed a generous collaboration with us both technically and materially in all aspects of *in vitro* transcription with T7 RNA polymerase; W. Konigsberg and B. Alberts for purified T4 DNA polymerase; D. Irvine and D. Hartz for their enthusiastic and seminal discussions. Supported by NIH grants GM 28685 and GM 19963.

25 April 1990; accepted 21 June 1990