The T4 protein, RegA, is a translational repressor that blocks ribosome binding to multiple T4 messages by interacting with the mRNAs near their respective AUG start codons. Other than the AUG, there are no obvious similarities between the affected mRNAs. High-affinity RNA ligands to RegA were isolated using SELEX (systematic evolution of ligands by exponential enrichment). The selected RNAs exhibited the consensus sequence 5′-AAAAUGUAUGUA-3′. The AUG was invariant, suggesting that it is the primary effector of binding specificity. The UU immediately 5′ to the AUG and the upstream poly(A) tract were highly conserved among the selected RNAs. Boundary and footprinting experiments are consistent with the consensus sequence defining the RegA-binding site. Interestingly, chemical modification and nuclease digestion data indicate that the RNA-binding site is single-stranded, as if RegA discriminates between targets based on their primary sequence, not their secondary structure. Minor variations from the consensus at positions other than the universally conserved AUG have little effect on RegA binding, but accumulation of mutations has a profound effect on the interaction. Comparison of the in vivo targets for RegA to the SELEX-generated consensus suggests a repression pattern whereby the translation of individual messages is sequentially halted until the least similarly affected message, the regA gene itself, is repressed.

Translational regulation has been shown to be an important means for controlling gene expression in a variety of organisms, both prokaryotic (1, 2) and eukaryotic (3, 4). One of the more interesting regulatory mechanisms involves the repression of translation caused by RNA-binding proteins interacting specifically with mRNAs. Many of these translational repressors function by directly competing for mRNA binding with ribosomes, thus decreasing the level of translational initiation (5). Among the well-characterized repressors, the bacteriophage T4 translational repressor, RegA, is unusual in that it affects the translation of many independent messages.

The expression of at least nine T4 genes is reduced in the latter stages of the phage life cycle by the autoregulated product of the regA gene (6). Transcription of these genes is not altered, and thus RegA-mediated repression occurs post-translationally (7). Genetic analysis and in vitro footprinting indicate that RegA specifically interacts with several of the regulated mRNAs near their translational start sites (8–10). The presence of RegA alters the binding of the 30 S subunit of Escherichia coli ribosomes to these mRNAs, thus preventing translational initiation (9). Taken together, these data are consistent with RegA altering gene expression in vivo by obstructing ribosome binding to specific mRNAs.

Although RegA repression is specific, the mRNA sequences that are bound by the repressor display few similarities. A consensus for the region surrounding the translational start sites of the affected mRNAs is indistinguishable from one generated for all of the known T4 messages (11). The lack of a distinctive consensus for the RegA-bound RNAs is not terribly surprising given that the putative repressor-binding site lies within the RNA domain used for translational initiation. Elements such as the AUG start codon and Shine-Dalgarno sequence are apparent in all of the messages of T4; thus repressed and unrepressed messages will necessarily share these characteristics within the RegA binding domain. However, it has been shown that single-site substitutions in this region of the affected messages can reduce RegA binding by more than 2 orders of magnitude (12). In addition, NMR studies of an RNA fragment harboring a G → U substitution in the same region indicate that both the native and mutated RNAs have similar single-stranded conformations (13). Results from mutational (14) and deletion analyses (15) suggest that the C terminus of T4 RegA protein provides the nonspecific nucleic acid binding component but the ability to discriminate between sequences resides elsewhere on the protein.

An understanding of the RNA elements required for binding by RegA would best be achieved by separating repressor binding from the in vivo requirements for translation. We used SELEX (16), an in vitro method for isolating RNAs from a random sequence population that has the highest affinity for a target protein, to identify the RNA components required for RegA binding without the requisite need for translation. Fifteen rounds of selection yielded a clear consensus. Surprisingly, the consensus was not a structural motif as is generally the case for RNA-binding proteins; rather, the consensus was a specific sequence. Results reported here led us to propose that RegA recognizes its targets in a sequence-preferred structure-independent manner.

**EXPERIMENTAL PROCEDURES**

**Protein Purification**—The RegA protein was purified as described (17).

---

* This work was supported by National Institutes of Health Grants GM53651 (to C. K.), GM19963 (to L. G.), and AI39186 (to P. A.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Molecular, Cellular, and Developmental Biology, University of Colorado, Campus Box 347, Boulder, Colorado 80309.

1 The abbreviations used are: SELEX, systematic evolution of ligands by exponential enrichment; PCR, polymerase chain reaction; CMCT, 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate; kethoxal, 2-keto-3-ethoxy-N-butyraldehyde.
Selection of RNA Ligands for RegA—A nucleic acid library possessing 5' and 3' fixed regions surrounding a 30-nucleotide randomized region was generated as described (18). 10^12 RNA molecules comprising approximately 10^14 unique sequences were incubated in 100 μl of RegA buffer (10 mM Hepes, pH 7.2, 100 mM NaCl, 5 mM MgCl2, and 0.01 mM dithiothreitol) with 10 μM RegA for 5 min at 25 °C. The binding reactions were applied to nitrocellulose filters, which preferentially retain RNAs that are bound to protein, and the filters were washed with 10 ml of RegA buffer. The protein-bound RNA was extracted from the protein/filter as described (19). The RNA was reverse-transcribed using the 3' primer SG1 (GCC GGA TCC GGCT CAT GTC GAA), and PCR amplification was carried out using both SG1 and the 5' primer SG1 (CCG AAG CTT ATT ACG ACT CAT TAT AGG GAC TCA ATA AAC GCT CAA). The resulting PCR product was transcribed with T7 RNA polymerase (20). The RNA was gel-purified (19) and used in the subsequent round of RegA binding. This process was continued with decreasing amounts of RegA (to increase selection stringency) for 15 cycles. The round 15 PCR product was restricted with HindIII and BamHI and ligated into similar sites of pUC 18, and the resulting plasmids were used to transform DH5α as described (21). Clonal inserts were sequenced using standard methods.

RNA Ligand-RegA Binding Affinities—Dissociation constants for the interactions between RegA and various ligands were determined using an electrophoretic mobility shift assay (22). 50 pmol RNA that had been 5' end-labeled by T4 polynucleotide kinase was incubated with various concentrations of RegA (0.1 nM–1 μM) in 10 μl of RegA buffer at 25 °C for 5 min. The bound and unbound RNAs were separated by electrophoresis through a non-denaturing 8% polyacrylamide gel. For the quantitative analysis shown in Table I, the relative amounts of bound and unbound RNA in each lane were quantified by scintillation counting of appropriate bands.

Boundary Analysis—The minimal 5' sequence required for the binding of RegA was determined basically as described (23). Five picomoles of partially hydrolyzed 5' end-labeled RNA was incubated with 5 or 15 nM RegA in 500 μl of RegA buffer at 25 °C for 5 min. The RegA-bound RNA fragments were separated from the unbound fragments by nitrocellulose filter binding. Bound RNAs were extracted as above and sequenced by polyacrylamide gel electrophoresis. Partial RNase T1 digests were performed as described (23).

RNA Structure Analysis and Protein Footprinting—RNAs were chemically modified using 1-cyclohexyl-3-(2-morpholinophenoxy)carboxy-imide metho-p-toluenesulfonate CMCT, kethoxal (2-keto-3-ethoxy-N-butylcarboxyde), and dimethyl sulfate as described (24) with a few exceptions. Each modification reaction contained 15 pmol of RNA in 50 μl. Each modification reaction was stopped and reverse-transcribed as described (25). In addition, RNAs were partially digested with either ribonuclease S1 (0.2 units/reaction), T1 (0.2 units/reaction), or ribonuclease A (0.1 μg/ml). Enzymatic digestions were carried out with 4 pmol of RNA/10-μl reaction at 25 °C for 1 min. For footprinting experiments, RegA and RNA were incubated together in binding buffer at 25 °C for 7 min before adding RNase for 1 min.

Generation of RNAs with 3' Disruptions—An oligonucleotide of sequence 5'-CCGGGGCTTTTGTGCAATT-3' was used in PCR reactions along with the 5' primer used in the selection and DNA preparations of the various clones to provide a template whose transcription product possessed a disrupted 3'-RegA-binding site. The RNAs were internally labeled by including [α-32P]UTP in the transcription reactions.

RESULTS

Translational repression of several early T4 genes results from the binding of the regA gene product to specific mRNAs near their respective AUG start codons, preventing the initiation of translation (9). Although the sequences and putative RegA-binding sites of several of these mRNAs are known, a shared primary or secondary structure is not obvious (11). SELEX was used to uncover the binding site specificity of RegA. Five cycles of selection with 10 μM RegA for 5 min at 25 °C were used to obtain dissociation constant for RegA of approximately 5 nM (Fig. 2, A and C). At RegA concentrations of approximately 20 nM, a second, slower migrating band was apparent (Fig. 2, A and C). All of the label shifted to this region of the gel at the highest concentration of protein. A simple explanation for these results is that a single binding event occurs at lower RegA concentrations and that an additional protein binds to each ligand at higher concentrations yielding the second shift. This is consistent with the sequence analysis above that suggests that two RegA-binding sites exist on each of these RNAs.

The AUG of the putative second binding site was altered to UUU, and the 3' most nucleotides were removed to test whether a single binding event could be observed. The modified ligands produced only a single shift (Fig. 2, B and D) and displayed affinities for RegA that were the same as the full-length molecules from which they were derived. The findings are consistent with there being two independent RegA-binding sites associated with each of the two ligands. The 5'-binding sites of the two RNAs are apparently the higher affinity sites, with dissociation constants for RegA of approximately 5 nM.

Although the sequence data suggest a relative size of the RegA-binding site based on the region of greatest homology, a physical measure of the minimal domain required for high affinity binding to the protein was acquired using the boundary structures what we believe is the AUG start codon observed in all of the native mRNA targets of RegA. The 3' consensus of AAA is interesting when the 3' fixed region of UUCGACAUU is taken into account. The resulting sequence of AAAAAUUCCGCAUG compares favorably with the 5' consensus of AAAAAUUGUUAGUUA, where the first two As in the latter sequence are provided by the 5' fixed region. The similarities between the two independent sites within the selected RNAs suggest the existence of two binding sites for RegA on each molecule. Because the putative 5'-binding site is made up almost entirely of nucleotides that were selected from the variable region, the relative conservation of each of the positions is quite telling as to the nature of the RegA binding interaction. The AUG is absolutely conserved among the ligands, suggesting that it is the primary effector of binding affinity. The poly(A) tract is also highly conserved, with its apparent optimum position being five nucleotides upstream of the AUG. The UU immediately following the upstream poly(A) tract and the AA following the AUG are also present in >90% of the ligands. The remaining positions are less conserved, although the level of conservation of these sequences within the putative binding domain is still quite high (>75%).

Binding affinities between RegA and specific ligands were measured using a gel mobility shift assay. The two sequences that occurred in multiple clones were PCR-amplified from plasmid DNA, transcribed, and radiolabeled. These two ligands were incubated with a range of RegA concentrations, and the bound and unbound RNAs were separated by non-denaturing polyacrylamide gel electrophoresis. Protein binding, as witnessed by a mobility shift, was first observed in the low nM range for both ligands, with their dissociation constants occurring at approximately 5 nM (Fig. 2, A and C). At RegA concentrations of approximately 20 nM, a third, slower migrating band was apparent (Fig. 2, A and C). All of the label shifted to this region of the gel at the highest concentration of protein. A simple explanation for these results is that a single binding event occurs at lower RegA concentrations and that an additional protein binds to each ligand at higher concentrations yielding the second shift. This is consistent with the sequence analysis above that suggests that two RegA-binding sites exist on each of these RNAs.

5'-AAAAAUUGUUAUGUUAUUGUGGUGUGUGUAAA-3'

Fig. 1. Consensus sequence from RegA SELEX. The two nearest As at the 5' end represent the last two nucleotides at the 3' end of the 5' fixed region. Our proposed RegA-binding site is in bold. A total of 25 isolates were sequenced. Below each position is the frequency at which that position varies and the base substitution. The two As from the 5' fixed and the AUG are invariable.
Each of the four RNAs described in the binding experiments above were 5' end-labeled and subjected to partial alkaline hydrolysis. The resulting RNA fragments were bound to RegA and passed through nitrocellulose filters. The fragments that still possessed the protein-binding site were preferentially retained on the filters. The recovered RNA fragments were characterized via polyacrylamide gel electrophoresis and autoradiography (Fig. 3).

The two full-length RNAs yielded two distinctive boundaries, the more efficiently retained occurring up to the AUG of the putative 3'-RegA-binding site and a second boundary occurring at the 3' end of the UAA of the putative 5' binding site. These data are consistent with there being two independent binding sites. The presence of two binding sites could either increase the relative affinity of the RNA for protein or increase the efficiency of filter retention by increasing the amount of protein bound per RNA, thus enhancing the percentage of fragments being retained during partitioning. There is binding after the 3' site is disrupted so long as the 5' site is maintained, but once the fragments lose nucleotides at the 3' end of the proposed 5'-binding site, filter retention is lost altogether. The RNAs with the 3' site disruptions displayed only a single boundary at the 3' end of the predicted 5'-binding site (Fig. 3).

Partial nuclease digests of the full-length and 3' site-disrupted versions of ligand A indicated that most of the nucleotides were sensitive to the single-strand-specific RNases (Fig. 4). The addition of two concentrations of RegA to the various RNA/RNase reactions resulted in the protection of specific bases from nuclease attack. Complete protection of the bases between positions A22 and A35 of both ligands was observed at 100 and 500 nM RegA (Figs. 4A and 3B). These RegA-induced protections span the putative 5'-binding sites of the two RNAs. Bases between A35 and U42 show either equivalent or enhanced nuclease activity in the presence of RegA. Most of the bases
between positions G44 and U59 in the full-length version of ligand A are completely protected from nuclease attack by both concentrations of repressor (Fig. 4A). This second RegA footprint covers the putative 3′-binding site of the RNA. Except for positions U48 and U58, which are partially protected by 500 nM RegA, the corresponding bases in the 3′ site-disrupted version of ligand A are accessible to the nucleases in the presence or absence of RegA (Fig. 4B). The loss of the footprint at the 3′ end of the second ligand is once again consistent with the assertion that two RegA binding events occur on the full-length ligands and that binding at the 3′ site is dependent on there being an AUG.

The apparent lack of RNA secondary structure observed in the nuclease digestions was investigated further using several single-strand-specific base-modifying reagents. The nucleotides making up the putative RegA-binding sites for ligands A and B were completely sensitive to the reagents (Figs. 5A and 4B). In fact, very few of the bases in the entire RNAs escaped modification, indicating that the ligands were devoid of structures that rely on Watson-Crick base pairing. These data suggest that RegA interacts with its RNA-binding sites in a structure-independent manner, a property that is unlike the well-characterized RNA-binding proteins.

Because none of the mRNAs affected by RegA in vivo possess the SELEX-generated consensus sequence, it was of interest to understand the relative effect that alterations from the consensus have on RegA binding affinity. Using RNAs whose 3′-binding site had been disrupted as above, the consensus sequence and several variants were tested for RegA binding. As seen in Table I, the UUGUU region 5′ to the AUG and the UAA 3′ to the AUG can undergo single base changes with little effect on binding. Thus the binding data are consistent with the sequence data that indicated that these two domains are important, but not essential, for RegA binding. Mutations in the AUG probably have a much greater effect as witnessed by the apparent loss of RegA binding caused by mutating the AUG of the 3′-binding site to UUU (Fig. 2). In contrast to the slight effect that single base changes have on RegA binding, multiple changes cause more significant decreases in affinity (Table I). This suggests that multiple mutations in this conserved region may have an additive effect on the interaction, which can have a profound consequence on the binding affinity of the site.

**DISCUSSION**

The T4-encoded RegA is one of the few known proteins that regulates the translation of multiple transcripts (26). The mechanism by which the protein discriminates between messages has remained a mystery, as the makeup of the start sites of the affected mRNAs are not statistically unlike those that are unaffected (11). Using the SELEX protocol, a set of RNAs were generated that possessed high affinity for RegA. Characterization of these ligands indicates that the RegA consensus binding site is AAAAUUGUUAUGUAA. Stable secondary structures that rely on canonical base pairs do not exist for the consensus, suggesting that RegA discriminates between RNAs based on primary sequence rather than secondary structure. The absence of Watson-Crick base pairing in the RNA binding domain of the SELEX ligands is supported by chemical modification and nuclease digestion results. A lack of secondary structure potential has likewise been observed in the in vivo
binding sites for the protein (27). The consensus sequence suggests that the relative site of interaction between RegA and the affected mRNAs includes the first two codons of the message plus the nine nucleotides immediately upstream.

The sequences of the T4 mRNAs that are known to be regulated by RegA were compared with the SELEX-generated consensus using the AUG start codon for alignment. In addition to the AUG, most of the RNAs possess a UU 5' to the translational start site and a poly(A) tract between two to eight nucleotides upstream (Fig. 6). Although these features are not at fixed distances from the AUG, their presence is suggestive of similar interactions between substrate and repressor. Interestingly, the mRNAs display varying degrees of similarity to the consensus binding sites, with the regA gene itself being the least similar. If the binding affinity of RegA for the mRNAs correlates with their similarity to the consensus, then the translation of the various genes would be halted successively until the regA gene was repressed. All mRNAs that were less similar to the consensus than the regA gene would not be repressed, as the concentration of RegA would be held below a threshold level that was a function of the binding affinity of the RegA-binding site of the regA gene.

The SELEX-generated consensus sequence has a stop codon following the AUG start in the RegA-binding site; thus, the highest affinity site for RegA could not exist within an mRNA that encodes a polypeptide. Why would a translational repressor be selected with such an RNA-binding site? A possible explanation could be that mRNA binding is actually a secondary function and that RegA is optimized to bind a cellular RNA. The T4 genome was searched for sequences that match the consensus. No exact matches of the consensus RegA-binding site were found, but several sites with properly aligned sequences (upstream poly(A), U/G tract, AUG, and UAA) were uncovered. One of the most similar sequences was AAAAUAUAUGUA, which is located within the dam gene (28). This region of the T4 genome has been heavily studied because it supports T4-dependent replication. Current opinion holds that the region is transcribed by the host RNA polymerase, giving rise to an RNA that primes T4 replication. If a transcript is produced from the region of the T4 genome, then the high affinity site for RegA actually exists in vivo during a time that RegA is produced (29, 30). It is postulated that RNA-protein and protein-protein interactions involving RegA localize the various components of the T4 replisome to origins of replication. The presence of a potentially high affinity RegA-binding site within a T4 genomic domain associated with a replication origin is intriguing, as it provides a possible mechanism for localizing replisome assembly within the cell to a region where replication is initiated.

Acknowledgments—We thank NeXstar Pharmaceuticals, Inc. and the Gold Laboratory for continued support and K. Handwerger and E. Spicer for valuable discussions and reading of this manuscript.

REFERENCES
1. Gold, L. (1988) Annu. Rev. Biochem. 57, 199–233
2. McCarthy, J. E., and Gualerzi, C. (1990) Trends Genet. 6, 78–85
3. Konak, M. (1992) Annu. Rev. Cell Biol. 7, 197–235
4. Standart, N., and Jackson, R. J. (1994) Biochimie (Paris) 76, 867–879
5. Campbell, K. M., Stormo, G. D., and Gold, L. (1983) in Gene Function in Prokaryotes (Beckwith, J., Davies, J., and Gallant, J. A., eds) pp. 185–210, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
6. Wiberg, J. S., and Karam, J. D. (1983) in The Bacteriophage T4 (Mathews, C. K., Kutter, E. M., Mosig, G., and Berget, P. B., eds) pp. 193–201, American Society for Microbiology, Washington, D. C.
Base-sequence Recognition and Binding to RNA by RegA

7. Trimble, R. B., and Maley, F. (1976) J. Virol. 17, 538–549
8. Karam, J., Gold, L., Singer, B. S., and Dawson, M. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 4669–4673
9. Wintzer, R. B., Morrissey, L., Gauss, P., Gold, L., Hsu, T., and Karam, J. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 7822–7826
10. Webster, K. R., Adari, H. Y., and Spicer, E. K. (1989) Nucleic Acids Res. 17, 10547–10556
11. Unnithan, S., Green, L., Morrissey, L., Binkley, J., Singer, B., Karam, J., and Gold, L. (1990) Nucleic Acids Res. 18, 7083–7092
12. Webster, K. R., and Spicer, E. K. (1990) J. Biol. Chem. 265, 19007–19014
13. Szewczak, A. A., Webster, K. R., Spicer, E. K., and Moore, P. B. (1991) J. Biol. Chem. 266, 17832–17837
14. O’Malley, S. M., Sattar, A. K. M., Williams, K. R., and Spicer, E. K. (1995) J. Biol. Chem. 270, 5107–5114
15. Webster, K. R., Keill, S., Konigsberg, W., Williams, K. R., and Spicer, E. K. (1992) J. Biol. Chem. 267, 26097–26103
16. Tuerk, C., and Gold, L. (1990) Science 249, 505–510
17. Kang, C., Chan, R., Berger, I., Lockshin, C., Green, L., Gold, L., and Rich, A. (1995) Science 268, 1170–1173
18. Brown, D., and Gold, L. (1995) Biochemistry 34, 14765–14774
19. Chen, H., Brown, D., and Gold, L. (1997) Methods Enzymol. 275, in press
20. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Nucleic Acids Res. 15, 8783–8788
21. Maniatis, T., Fritsch, E. P., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
22. Carey, J. (1991) Methods Enzymol. 206, 103–117
23. Tuerk, C., Eddy, S., Parma, D., and Gold, L. (1990) J. Mol. Biol. 213, 749–761
24. Stern, S., Moazed, D., and Noller, H. F. (1988) Methods Enzymol. 164, 481–489
25. Binkley, J., Allen, P., Brown, D., Green, L., Tuerk, C., and Gold, L. (1995) Nucleic Acids Res. 16, 3198–3205
26. Miller, E. S., Karam, J. D., and Spicer, E. K. (1994) in Molecular Biology of Bacteriophage T4 (Karam, J., ed) pp. 193–295, American Society for Microbiology, Washington, D. C.
27. Miller, E. S., Karam, J., Dawson, M., Trojanowska, M., Gauss, P., and Gold, L. (1987) J. Mol. Biol. 198, 397–410
28. Macdonald, P. M., and Mosig, G. (1984) EMBO J. 3, 2863–2871
29. Cardillo, T. S., Landry, E. F., and Wiberg, J. S. (1979) J. Virol. 32, 905–916
30. Campbell, K., and Gold, L. (1982) in Interaction of Translational and Transcriptional Controls in the Regulation of Gene Expression (Grunberg-Manago, M., and Safer, B., eds) pp. 69–81, Elsevier Science Publishing Co., Inc., New York
Single-stranded RNA Recognition by the Bacteriophage T4 Translational Repressor, RegA

David Brown, Joelle Brown, ChulHee Kang, Larry Gold and Patrick Allen

*J. Biol. Chem.* 1997, 272:14969-14974.
doi: 10.1074/jbc.272.23.14969

Access the most updated version of this article at [http://www.jbc.org/content/272/23/14969](http://www.jbc.org/content/272/23/14969)

Alerts:
- When this article is cited
- When a correction for this article is posted

[Click here](http://www.jbc.org/) to choose from all of JBC's e-mail alerts

This article cites 24 references, 10 of which can be accessed free at [http://www.jbc.org/content/272/23/14969.full.html#ref-list-1](http://www.jbc.org/content/272/23/14969.full.html#ref-list-1)