RIBOSOME BINDING SITE SEQUENCES AND FUNCTION

John Childs, Keith Villanueva, Doug Barrick, Thomas D. Schneider, Gary D. Stormo, Larry Gold, Moshe Leitner, and Marvin Caruthers

1Radiation Biology Branch, Atomic Energy of Canada Ltd., Chalk River, Ontario K0J 1RO

2Department of Molecular, Cellular and Developmental Biology; 3Department of Chemistry, University of Colorado, Boulder, Colorado 80309

4Present address: Israel Institute for Biological Research, Ness Ziona, Israel.

ABSTRACT In order to determine the contribution of every base within an Escherichia coli ribosome binding site, we have constructed a plasmid for cloning random sequence, synthetic single-stranded DNA at the 5' end of a lacZ gene that has a promoter but lacks both the first 8 non-essential codons and a ribosome binding site. When this vehicle is cut between lacZ and the promoter with BglII and PstI, 5' and 3' overhangs are created. We have successfully cloned single-stranded DNA with ends complementary to those of BglII and PstI into this vector, a procedure we term "notch cloning". We have cloned the ribosome binding site of rpoA, the alpha subunit of RNA polymerase of E. coli, with random incorporation of nucleotides at positions -3 and 0, where 0 is the first base of the initiation codon, and have determined beta-galactosidase levels for several of the sequences. Additionally, we have cloned completely random sequences from -11 to 0 to determine their effects on ribosome binding.
INTRODUCTION

Despite the considerable amount of attention that has been paid to the mechanism of ribosome binding to messenger RNA in Escherichia coli, the precise details remain elusive. Although the strength of ribosome binding can vary by at least 870-fold (9,10), it is not possible to predict the activity from the sequence. The elements involved in binding that have been studied in greatest detail are the initiation codon and the Shine and Dalgarno region (13). These two regions are usually separated by a spacer of 5 to 9 bases (15).

It is apparent, however, that these elements alone do not define a binding site. Shine and Dalgarno-like sequences followed at the appropriate interval by an AUG codon are found both at the start of genes and within genes. Stormo et al. (15) examined these sequences and showed that true gene starts, unlike non-initiating AUG's, have a non-random distribution of bases between -20 and +13 in addition to the Shine and Dalgarno region and initiation codon. As this is the region protected by ribosomes in bind-and-chew experiments (5) it is likely that the non-randomness is related to ribosome binding.

In order to further study ribosome binding sites we have constructed a plasmid that permits us to clone single-stranded DNA into the 5' end of a \( \text{lacZ} \) gene that has a promoter but no ribosome binding site. This allows us to clone synthetic ribosome binding sites with random changes at any number of positions 5' to the second base of the initiation codon. Our eventual aim is to determine the influence of each individual position on ribosome binding and thus from the sequence predict initiation rates. We report here our first constructions.

MATERIALS AND METHODS

(a) Single-Stranded DNA

Single-stranded DNA was synthesized on an Applied Biosystems Model 380A DNA Synthesizer.

(b) Bacteria and plasmids

\( E. \text{coli} \) 79-02 (C600 hsdR, hsdM, thr, leu, str, \( \Delta \text{laczY-pro} \), F' \( \text{traD36, lacIq, lacZ AM15, pro}^+ \)).
Plasmids used to construct pBC26 and pBC29 (figure 1) were kindly donated by M. Casadaban, L. Klobutcher, M. Russel, and M. Mockensturm.

(c) DNA Sequencing

pBC26 and related plasmids contain the origins of replication of both colE1 and phage fl, and so may be sequenced by the dideoxy method (11) using single-stranded DNA obtained by the method of Dente et al. (3). Plasmid-containing E. coli 79-02 strains were grown overnight in H-broth (14) plus 4 mg/l vitamin B1 and 25 mg/l ampicillin (amp), were diluted 50-fold in L broth plus amp and then infected at an MOI of >20 with phage fl (IR1) (4). The cultures were aerated gently for 1 hour and then vigorously for 5 hours or overnight. Bacterial cells were removed by centrifugation and phage particles were precipitated with polyethylene glycol 6000. Single-stranded DNA was extracted from the particles with phenol and then sequenced using either the universal M13 primer (16) or our own synthetic primer which hybridizes downstream of the M13 primer.

(d) Beta-galactosidase Assays

We used kinetic rate measurements of o-nitrophenyl-beta-D-galactoside hydrolysis (ONPG; 7); the method will be described elsewhere (Barrick et al., in preparation).

RESULTS

(a) Design of Single-Stranded Inserts

Single-stranded DNA to be cloned into pBC29 (figures 1 and 2) must have BgIII and PstI complementary ends (figure 3). One synthetic DNA (called Des11) has two variable nucleotides (all 4 bases inserted during synthesis) and was used to test the system. With A's at the variable positions -3 and 0, Des11 is the same as the ribosome binding site of rpoA (RNA polymerase alpha subunit; 8) from positions -15 to +6. Two other DNA's, Des12 and Des19 (figure 3), each contain twelve variable nucleotides. These DNA's are designed to ask how nucleotides 5' to an initiation codon can influence translation.
FIGURE 1. Structure of pBC26 and pBC29. The sources of DNA for pBC26 (defined by restriction cuts) are:

**EcoRI to HindIII**  Pra2 promoter and lac operator shown in figure 2 (1). (size 79 bp).

**HindIII to BamHI**  Cloning sites shown in figure 2 (18 bp).

**BamHI to AvaI**  The lac fragment, derived from pMC1403 (2), starts at the 8th codon of lacZ and ends at the AvaI site in lacY (size 3313 bp).

**AvaI to SalI**  This DNA is from AvaI to EcoRI of pBR322 with mutant PstI and HincII sites from pUC8 (16). These mutations do not affect the amp resistance of the plasmid. The EcoRI site from pBR322 was filled in with the Klenow fragment of DNA polymerase I and then SalI linkers were added (total size 2940 bp).

**SalI to EcoRI**  This is DNA from nucleotides 5148 to 6407 of f1 phage obtained from pD4 (4) as an EcoRI fragment. One EcoRI site was changed to SalI as described above (size 1274 bp). The sites EcoRI, HindIII, BglII, PstI, BamHI, and SalI are unique in pBC26.

pBC29 has lambda DNA from nucleotides 35712 to 37005 (12) inserted between the BglII and PstI sites of pBC26.
Ribosome Binding Sites

![Diagram of DNA sequence and binding sites](image)

FIGURE 2. DNA sequence of pBC26 from the end of the f1 DNA to the start of the lacZ DNA.

**Des11**

```
AGATCAAGAGAGAGACNCANTGCA
```

**Des12**

```
5' GATCAANNNNNNNNNNTGCA 3'
```

**Des19**

```
5' GATCAAAARRRRRRRRNNNTGCA 3'
```

FIGURE 3. Sequences of Des11, Des12 and Des19 single-stranded DNAs. N indicates that all four nucleotides were inserted during synthesis, while R indicates that only purines were inserted. All three DNAs can be cloned into the notch between the BglII and PstI sites of pBC26 as shown for Des11. Bases 0, 1 and 2 form the initiation codon in frame with the lacZ gene to the right. The sequence of the rpoA ribosome binding site from -17 to +6 differs from that of Des11 clones at only positions -16, -3 and 0, which in rpoA are C, A and A respectively.
(b) Notch Cloning

Plasmid pBC29 was cut with BglII and the 5' phosphates were removed with alkaline phosphatase. The DNA was extracted with phenol and then cut with PstI. Single-stranded DNA was kinased at the 5' end; the cut vehicle provides the phosphate required to ligate the 3' end (figure 3). The single-stranded fragments were then mixed at a 10-fold molar excess with the vector at a final total DNA concentration of 4 µg/ml in ligase buffer, incubated at 65°C and then allowed to cool slowly for 1 hour to about 30°C. An equal volume of ligase buffer containing 4 units/µl of T4 DNA ligase was added and the mixture incubated at 16°C for 2 hours or overnight. Following transformation of 79-02, similar numbers of both white and blue colonies were obtained on plates containing ampicillin and the indicator substrate 5-bromo-4-chloro-3-indolyl-beta-D-galactoside (Xgal). We obtained about $10^4$ transformants per µg of vector DNA. Plasmids obtained from these transformants were first screened for the presence of restriction sites in the vicinity of the notch. Plasmid DNA isolated from the white colonies did not have HindIII, BglII, PstI or BamHI sites, indicating that they were probably deletion mutations. Plasmids from the blue colonies fell into three classes. The first, which were from extremely pale blue colonies, were cut by all 4 enzymes and were about 8.9 kb, indicating that they were parental pBC29. The second were not cut by any of these enzymes and were probably deletions, while the third were cut by HindIII and BamHI, but not by BglII and only in a minority of cases by PstI. As any of the single-stranded inserts should destroy the BglII site and 3/4 of the PstI sites, it was anticipated that these would have the inserted DNA. Subsequent sequencing confirmed that all of these had inserts. More than 80% of all the blue colonies had plasmids with Des inserts.

(c) Cloning DesII

Out of the 16 possible DesII sequences, 5 were obtained. We synthesized and notch cloned the rpoA
wild-type sequence separately. The sequences and their corresponding beta-galactosidase values (in parenthesis) normalized to the wild-type rpoA sequence were:

\[
\begin{align*}
&5' \text{gatcaaagagaggacAcaAtgca (100)} \\
&\downarrow \quad \downarrow \\
&\text{CcaT} (2.8) \quad \text{CcaA} \quad (62) \\
&\downarrow \\
&\text{CcaC} (2.5) \quad \text{TcaA} \quad (27) \\
&\downarrow \\
&\text{TcaG} \quad (12)
\end{align*}
\]

The highest level of beta-galactosidase was obtained with the rpoA sequence which has an ATG start codon and an A at -3. Lower levels were obtained from inserts with C or T at -3. A GTG instead of ATG had an even lower activity. The two other sequences, although giving rise to blue colonies, had barely detectable beta-galactosidase levels.

(d) Cloning Des12 and Des19

The single-stranded DNA Des12 (figure 3) was notched cloned. This DNA has 12 random positions, for which there are nearly 17 million possible clones. Des19 also has 12 random bases, but 4 of these have only purines in the Shine and Dalgarno region. We have sequenced and measured the beta-galactosidase values for 54 Des12 clones, and 8 Des 19 clones. In the random DNA, 239 bases were A, 140 were C, 137 were G and 228 were T. The di-nucleotide composition is well predicted by this mono-nucleotide composition, and there were no position effects on nucleotide incorporation during synthesis.

**DISCUSSION**

We have described the construction of a plasmid for the cloning of single-stranded DNA. Through the apposition of 5' and 3' overhangs we can insert single-stranded synthetic DNA that can subsequently be replicated
to give viable plasmids. The single-stranded DNA may be random at any number of positions, as long as the fixed ends of the molecule are complementary to the notched vehicle DNA. Other workers have randomized nucleotides as a step toward understanding a nucleic acid binding site. Our method allows randomization of long sequences, limited only by the quality of extended DNA synthesis.

Our first experiments proved that notch cloning works. We used a single-stranded DNA containing two variable positions in a sequence that resembles the ribosome binding site of the *E. coli* rpoA gene. From our first 13 Des11 inserts we obtained 5 of the 16 possible sequences. The wild-type rpoA sequence was not in this set; it was synthesized separately and notch cloned without difficulty. The data for beta-galactosidase expression from these six ribosome binding sites attest to the importance of the nucleotide at -3, as one might have predicted from the non-randomness at that position in collections of ribosome binding sites (15). Hui et al. (6) have also focused on this region and have randomized the -1 triplet. They found that these bases affect translational initiation rates. Our data also allow comparisons between the initiation codons AUG, GUG, UUG, and CUG. AUG directs translation at a somewhat higher level than GUG; UUG and CUG work poorly, at least within the context of the Des11 sequence.

We wish to fully understand ribosome binding sites, nucleotide by nucleotide. We thus turned to the single-stranded DNA molecules Des12 and Des19 (figure 3). At present 62 sequences have been determined, along with the corresponding levels of beta-galactosidase. An understanding of a nucleic acid site with twelve variable positions requires a mathematical analysis of perhaps 200 sequences (Stormo et al., in preparation). Thus, for twelve variable nucleotides, a relatively small number of isolates will lead to predicted translation levels for the whole set. While Des12 and Des19 clones probe the nucleotides that one conventionally assumes are responsible for translational initiation rates, our own data suggest that the nucleotides 3′ to the initiation codon are equally important (15). Thus another set of clones containing random synthetic sequences will probe that domain. The notch cloning vectors easily accommodate these experiments.
The concept of domain-directed randomization of DNA sequences seems a worthy addition to other forms of plasmid mutagenesis. This method should give powerful information about DNA, RNA, or protein structure and function. We are particularly intrigued by the use of domain-specific randomization of amino acid sequences spanning a residue at which a non-functional amino acid substitution has been found. Rare peptides that can substitute for another within a protein might easily be found.

ACKNOWLEDGEMENTS

Sidney Shinedling suggested the use of apposed 5' and 3' overhangs for single-stranded DNA cloning. We thank him for his contribution. This work was supported in part by NIH Grants GM28755 and GM28685.

REFERENCES


