The DNA sequence of the gene and genetic control sites for the excreted B. *subtilis* enzyme β-glucanase

N.Murphy[1], D.J.McConnell[1] and B.A.Cantwell[1,2]*

[1]Department of Genetics, Trinity College, Dublin 2, and [2]Arthur Guinness Son & Co. Ltd., St. James's Gate, Dublin 8, Ireland

ABSTRACT
     The sequence of a 1409 base pair restriction fragment con-
taining the B.subtilis gene for (1-3),(1-4)- β -D-glucan endoglu-
canase is reported. The gene is encoded in a 726 base pair seg-
ment. The deduced amino acid sequence of the protein has a hydro-
phobic signal peptide at the NH_2-terminus similar to those found
in five other secreted proteins from Bacillus. The gene is pre-
ceded by a sequence resembling promoters for the vegetative B.
subtilis RNA polymerase. This is followed by a sequence resemb-
ling a B.subtilis ribosome binding site nine nucleotides before
the first codon of the gene. Two sequences, one before and one
after the gene, can be arranged in secondary structures similar
to transcriptional terminators. There is also a short open read-
ing frame coding for a hydrophobic protein on the minus strand
just upstream from the β-glucanase gene. A possible role for
this gene in the control of expression of β-glucanase is suggested.

INTRODUCTION
     The genus Bacillus is the best characterised genus of gram-
positive organisms at the molecular genetic level.
     Information now available on the transcriptional and trans-
lational control sequences of Bacillus includes observations on
the molecular genetic control systems which operate during spor-
ulation (1,2) and in the course of infection by bacteriophage
(3,4,5,6). Promoter sequences have been characterised for the
σ^{55} - RNA polymerase of B.subtilis and some data are available
for promoter sequences of other forms of the holo-enzyme (2).
Complete nucleotide sequences are known for only a few Bacillus
genes. They include type 1 β-lactamase from B.cereus (7), peni-
cillinase from B.licheniformis (8) and α-amylases from B.sub-
tilis (9,10) and B.amyloliquefaciens (11). These are all extra-
cellular proteins and comparisons between the sequences of the
genes and the mature proteins have shown that they all possess a

hydrophobic signal peptide which is cleaved in the process of excretion (12,13).

We report here the complete nucleotide sequence of a 1409 bp DNA fragment containing the B.subtilis gene for the extracellular enzyme, (1-3),(1-4)- β-D-glucan endoglucanase. This gene has been cloned previously in E.coli (14).

MATERIALS AND METHODS

Bacterial strains and plasmids.

The plasmid pJG83, a derivative of pAT153 containing the β-glucanase gene from B.subtilis C120 (14) was used as a source of the β-glucanase gene.

The plasmid vector for subcloning was pUC8 or pUC9 grown in E.coli K-12 strain JM83 (15).

Media.

Bacteria were grown in L broth (10g Difco Bacto tryptone, 5g Difco yeast extract, 5g NaCl/litre adjusted to pH 7.2) and on L agar (L broth containing 20g/1 Oxoid No. 1 Agar). Antibiotics used in the selection of plasmids were added as appropriate to 10 µg/ml for tetracycline and 50 µg/ml for ampicillin.

Plasmid DNA isolation.

Plasmid DNA was isolated and purified from stationary phase cultures grown at 37° C using the rapid alkaline method of Birnboim and Doly (16), except that a phenol extraction step was included after the clearing spin.

Enzymes.

Restriction enzymes, T4 DNA ligase, Klenow fragment of DNA polymerase I and Bal 31 exonuclease were purchased from Bethesda Research Laboratories, New England Biolabs. or Boehringer Mannheim Gmbh. Each was used according to the manufacturers specifications.

Transformation.

Transformation of E.coli was performed according to Cohen et al. (17) except that the host cells were cultured at 37° C in L broth to A600 = 0.3 before treating with $CaCl_2$. After a heat pulse at 42° C for 2 mins. the cells were diluted 10-fold in L broth and incubated at 37° C for 2h to allow for expression before plating on L-agar plates containing the appropriate anti-

biotic. When selecting for recombinants of pUC8 the transform-
ation mix was plated on L-Amp plates containing a 3 ml overlay
of L-Amp agar with 5-bromo-4-chloro-3-indolyl-β-D-galactoside
(Xgal) as an indicator at 300 ug/ml.

DNA sequence analysis.

DNA fragments were 3' end-labelled with the aid of the Kle-
now fragment of DNA polymerase I with one labelled (α-^{32}P) dNTP.
Fragments cloned into pUC8 can be sequenced directly after cleav-
age with a second restriction enzyme (18). Otherwise labelled
fragments for sequencing were purified by electro-elution from
acrylamide gels. DNA sequence was determined according to the
procedures of Maxam and Gilbert (19). The products of the "G",
"G+A","T+C", "C" and "A>C" reactions were electrophoresed on 0.4
mm thick 8% or 20% polyacrylamide gels. Gels were dried onto
Whatman 3MM paper and exposed to Kodak X-AR5 film using Du-Pont
lightning plus screens at -70° C.

RESULTS AND DISCUSSION

(a)   Sequencing strategy.

The β-glucanase gene estimated to be approximately 710 base
pairs from the size of the β-glucanase protein (14,20) is known
to span the Sal I site in the plasmid pJG83 (14) containing the
β-glucanase gene as a 1.6kb EcoRI-PvuI fragment.

Sau 3A fragments from pJG83 containing Bacillus sequences
were isolated from acrylamide gels and subcloned into the Bam HI
site of pUC8. Ligation mixes were transformed into E.coli JM83
and clones were identified as white colonies on L amp plates con-
taining Xgal. To obtain the sequence from both DNA strands it
was necessary to generate clones in which the Sau 3A fragments
had been cloned in both orientations. The orientation of each
fragment was determined, where possible, by restriction patterns
with Alu I or Taq I. Plasmid DNA from clones where the orien-
tation of the Sau 3A fragment could not be determined was re-
stricted with EcoRI and HindIII, the fragment purified, and sub-
cloned into the EcoRI - HindIII site of pUC9.

All 7 Sau 3A fragments were thus sub-cloned in both orien-
tations and sequenced by the method of Ruther et al. (18). These
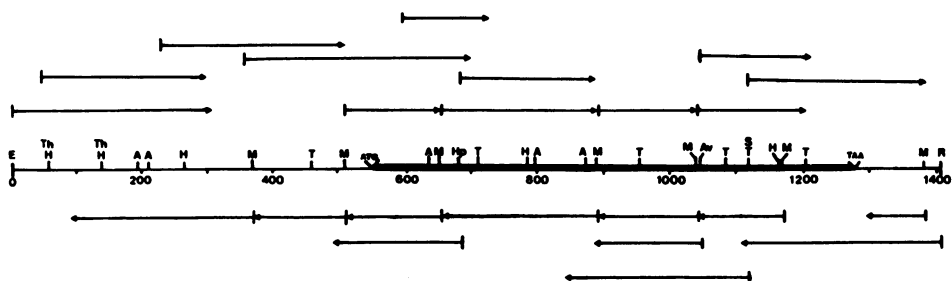fragments gave most of the sequence shown in Fig. 2. One region

FIGURE 1:   Restriction map and sequencing strategy of region surrounding the β-glucanase gene. The filled box indicates the coding sequence for the β-glucanase protein. The direction and extent of sequence determinations are shown by horizontal arrows. The sites of 3' end labelling are indicated by vertical lines on the arrows. The abbreviations of restriction endonuclease sites are: A,AluI; Av,AvaI; E,EcoRI; H,HhaI; Hp,HpaII; M,MboI; R,RsaI; T,TaqI; Th,ThaI.

at the left end of the EcoRI-RsaI 1.4 kb fragment (Fig. 1) could not be sequenced completely by this procedure because it is de-void of Sau 3A sites for about 370 base pairs. Accordingly, a series of deletions was constructed using Bal 31 to digest in from the EcoRI site, BamHI linkers were added and the resulting BamHI fragments were sub-cloned into the BamHI site of pUC8, and sequenced. In order to sequence across the rightmost six Sau 3A sites DNA fragments from pJG83 were end-labelled at the unique HpaII and SalI sites of the insert and sequenced as shown. The full sequence of 1409 base pairs is shown in Fig. 2. It was assembled using a computer programme to match sequences obtained from different runs. Most of the sequence (with the exception of short stretches at each end) has been determined for both strands (Fig. 1) and all restriction sites used for end-labelling have been sequenced across.

(b)   The location of the β-glucanase gene.

The sequence of Fig. 2 was searched for open reading frames sufficient to code for a protein of about M_r 25-26,000 (14,20). There is only one such sequence of 726 base pairs from base 554 to 1279. It spans the Sal I site known to lie in the β-glucanase gene (14) and has 242 codons equivalent to a protein of M_r 27,288. The N terminal amino acid sequence (amino acids 1-28) resembles the signal sequences of the precursor forms of five proteins of

Bacilli known to be processed during excretion when the signal
sequence is cleaved off (Fig. 3). Presumably the N terminal seq-
uence of the β-glucanase is also cleaved off when it is excreted.

Two segments have been distinguished in signal sequences of
proteins from organisms other than Bacilli (13, 21, 22).   There
is a short N terminal hydrophilic segment (2 - 11 amino acids)
often ending in an arginine or lysine followed by a longer (14 -
20) hydrophobic sequence. The cleavage site between the signal
sequence and the N terminal amino acid of the mature protein most
frequently follows the sequence ala-X-ala (13) although there is
considerable variation in the cleavage site sequence.

The five Bacillus signal sequences previously reported in-
clude four for which the cleavage site is known from comparisons
between the sequence of the precursor protein inferred from the
DNA sequence and the sequence of the mature protein (Fig. 3).
From these data it can be seen that Bacillus signal sequences
are similar to those from other species with hydrophilic segments
of 6 - 12 amino acids ending in a lysine or arginine and hydro-
phobic segments which are slightly longer (22 - 24) than the
range seen in other organisms. The four known cleavage sites all
follow an alanine residue. Between 60 and 75% of the residues in
the hydrophobic segments have a hydropathy index greater than 1
(23) as expected for peptides which are expected to interact with
membranes. There are striking homologies between the five signal
sequences shown by the boxed residues in Figure 3.

The putative β-glucanase signal sequence has similar pro-
perties to the other Bacillus signal sequences with a hydrophilic
segment of 6 amino acids ending at an arginine followed by a
hydrophobic segment which may end either at alanine 26 or alanine
28. Of the 22 amino acids from valine 7 to alanine 28, 16 have
hydropathy indices greater than 1. Moreover the β-glucanase
sequence shows homology with the other signal sequences as shown
in Figure 3. The sequence leucine-phenylalanine occurs at the
same position (13 - 14) relative to the end of the hydrophilic
N terminal segment in three of the six sequences including β-
glucanase and occurs four more times in the sequences at two
other positions 6-7 and 9-10. It is apparent from these consid-
erations that β-glucanase has a signal sequence typical of those

```
       GluPheAsnGluGluSerLeuHisTyrRyrArgPheValThrHisLeuLysPhePheAlaGlnArgLeuPheAsnGlyThrHisMetGluSerGlnAspAsp
   1   GAATTCAACGAAGAATCGCTTCACTATTATCGGTTCGTCACCCACTTAAAGTTTTTCGCGCAGCGTCTATTTAACGCCACTCACATGGAAAGCCAAGACCG


       PheLeuLeuAspThrValLysGluLysTyrHisArgAlaTyrGluCysThrLysLysIleGlnThrTyrIleGluArgGluTyrGluHisLysLeuThr
 101   ATTTTTTGCTGGATACAGTGAAAGAAAAGTATCATCGCGCGTATGAATGCACGAAGAAATCCAAACCTACATTGAGCGGGAGTATGAGCACAAGCTCAC


       SerAspGluLeuLeuTyrLeuThrIleHisIleGluArgValValLysGlnAla
 201   AAGTGACGAGCTGCTGTATTTAACCATTCACATAGAAAGGGTAGTTAAACAAGCATAATGAGAGCGCTGACATTTGTGTTTCCTTGTGTTCACTTTTTCT


 301   TACATTCACATATGAAAATGGTAGGATTGTTACTGATAAAGCAGGCAAAACCTAAATTGCAATGAGTGCCGGATCATCTCTGTCTGTGCTGATGGTAATTT
                                                                                                        AA

                                                                                          -35
 401   AGGTTTTTATTTTTTCAGAGGGAAGATGATGATAGTTACAGGATTCAAGTTAGTAAGATTCGATATTATCATTATTTTGACCGATGTTCCCTTTTGAAA
       TCCAAAAATAAAAAAGTCTCCCTTCTACTACTATCAATGTCCTAAGTTCAATCATTCTAAGCTATAATAGTAATAAAACTGGCTACAAGGGAAAACTTT
       ThrLysIleLysLysLeuProPheIleIleIleThrValProAsnLeuAsnThrLeuAsnSerIleIleMet                        S.D.

           -10                                      S.D.          f-MetProTyrLeuLysArgValLeuLeuLeuValThrGlyLeuPhe
 501   GAATCATGTAAGATCAACATAGAAAACGCTTTCAATGAAAGGGGAATGCCAATATGCCTTATCTGAAACGAGTCGTTGCTGCTTCTTGTCACTGGATTGTT
       CTTAGTACATTCTAGTTGTATCTTTTGCGAAAGTT
            -10                   -35


       MetSerLeuPheAlaValThrAlaThrAlaSerAlaLysThrGlyGlySerPhePheAspProPheAsnGlyTyrAsnSerGlyPheTrpGlnLysAla
 601   TATGAGTTTGTTTGCAGTCACTGCTACTGCCTCAGCTCAAACAGGTGGATCGTTTTTCGACCCTTTTAACGGCTATAACTCCGGTTTTTGGCAAAAAGCA


       AspGlyTyrSerAsnGlyAsnMetPheAsnGlyThrTrpArgAlaAsnAsnValSerMetThrSerLeuGlyGluMetArgLeuAlaLeuThrSerLeuAla
 701   GATGGTTATTCGAATGGAAATATGTTCAACTGCACGTGGCGGGCTAATAACGTATCAATGACGTCATTGGGTGAAATGCGTTTAGCCGCTAACAAGCCCAG


       TyrAsnLysPheAspCysGlyGluAsnArgSerValGlnThrTyrGlyTyrGlyLeuTyrGluValArgMetLysProAlaLysAsnThrGlyIleVal
 801   CTTATAACAAGTTTGACTGCGGGGGAAAACCGTTCTGTTCAAACATATGGCTATGGACTTTATGAAGTCAGAATGAAACCAGCTAAAAACACAGGGATCGT


       SerSerPhePheThrTyrThrGlyProThrAspGlyThrProTrpAspGluIleAspIleGluPheLeuLysAspThrThrLysValGlnPheAsn
 901   TTCATCGTTCTTCACTTACACAGGTCCAACAGATGGAACTCCTTGGGATGAGATTGATATCGAATTTTTAGGCAAAAGACACAACGAAGGTTCAATTTAAC


       TyrTyrThrAsnGlyAlaGlyAsnHisGluLysIleValAspLeuGlyPheAspAlaAlaAsnAlaTyrHisThrTyrAlaPheAspTrpGlnProAsnSer
1001   TATTATACAAATGGTGCAGGAAAACCATGAGAAGATTGTTGATCTCGGGTTTGATGCAGCCAATGCCTATCATACGTATGCCATTCGATTGGCAGCCAAACT


       IleLysTrpTyrValAspGlyGlnLeuLysHisThrAlaThrAsnGlnIleProThrThrLeuGlyLysIleMetMetAsnLeuTrpAsnGlyThrGly
1101   CTATCAAATGGTATGTCGACGGGCAATTAAAACATACTGCAACAAACCAAATTCCGACAACACTTGGAAAGATCATGATGAACTTGTGGAATGGCACGGG


       ValAspGluTrpLeuGlySerTyrAsnGlyValAsnProLeuTyrAlaHisTyrAspTrpValArgTyrThrLysLys
1201   TGTCGATGAATGGCTTGGCTCCTACAATGGTGTAAATCCGCTATACGCTCATTATGACTGGGTGCCGCTATAGAAAAAAATAATGCCCAAATGTCAAAGAAC
                                                                                    -----------→

                                   -35                     -10
1301   CTGCTGCAATATAGCAGGCTCTTATGATTGTAATGAGAAGTTGTTGGCCATGAATCTTCTATTCACTCATAATATCCTATTTGATCTTCTCCCCTCTGTAA
                                                          →

1401   ATCACGTAC
```

so far identified in Bacilli. By comparison with the other known
or postulated cleavage sites (Fig. 3) the cleavage of β-glucanase
possibly occurs to the right of alanine 26 or alanine 28 (13).

The codon usage in the β-glucanase gene is shown in Table 1.
Of 61 codons, 55 have been used. The unused codons were ACC
(thr 0/23), UGU (cys 0/2), AGG (arg 0/6), AUA (ile 0/7), CAC
(his 0/4) and CCC (pro 0/9). With the publication of the DNA
sequences of an increasing number of Bacillus genes it will be-
come possible to assess codon usage across the genome and invest-
igate differences between Gram-positive and Gram-negative organ-
isms (24).

(c) Transcriptional and translational signals.

There are several different forms of RNA polymerase in B.
subtilis which differ in σ factor and recognise different class-
es of promoter sequences (1,2,6). The main vegetative RNA poly-
merase has a σ factor of M_r = 55,000 and $σ^{55}$ promoters are the
best characterised in B.subtilis.

The consensus sequence of 9 $σ^{55}$ promoters is shown in Fig.
4. The 1409 base pair DNA sequence (Fig. 2) has been scanned
for sequences related to the -35 (TTGACA) and -10 (TATAAT) con-
sensus sequences which are usually separated by 17 or 18 base
pairs. There are three sequences (Fig. 4) which are in agree-
ment with the consensus and one of these is located 46 base pairs
from the ATG of the β-glucanase gene. It corresponds closely to
the consensus -35 and -10 sequences which are separated by 18
base pairs. Moreover 24 out of 30 of the base pairs upstream
from the -35 region are AT which is typical of $σ^{55}$ promoters (1).
Transcription usually starts at an A or G six or seven nucleo-
tides downstream from the -10 sequence which suggests that the
β-glucanase transcript starts at the A at 513 which is six nucleo-
tides distant from the -10 sequence.

FIGURE 2: DNA nucleotide and inferred amino acid sequence of
the B.subtilis β-glucanase gene and surrounding region. A
possible open reading frame of 86 codons 5' to the β-glucanase
gene is shown. The complementary DNA strand is included where
a possible short gene may be encoded (see text). Presumptive
"-35", "-10" and Shine-Dalgarno sequences are underlined.
Sequences containing inverted repeats both 5' and 3' to the β -
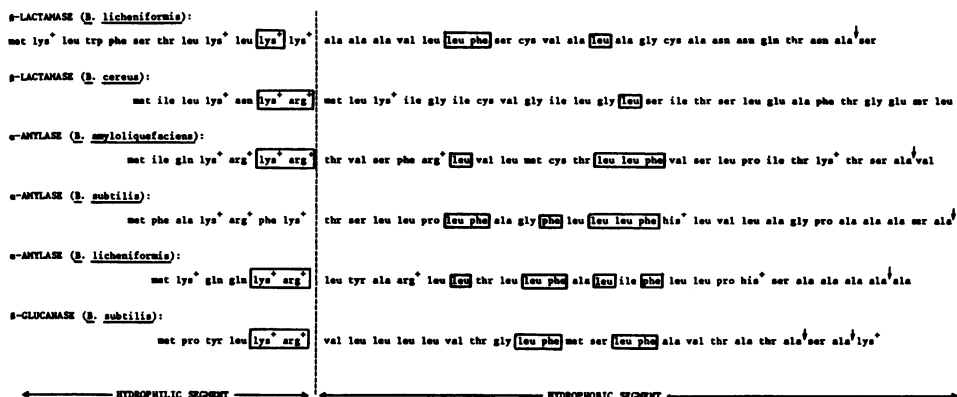glucanase gene are designated by —▶ ◀—

β-LACTAMASE (B. licheniformis):

met lys⁺ leu trp phe ser thr leu lys⁺ leu [lys⁺] lys⁺ ala ala ala val leu [leu phe] ser cys val ala [leu] ala gly cys ala asn asn gln thr asn ala↓ser

β-LACTAMASE (B. cereus):

met ile leu lys⁺ asn [lys⁺ arg⁺] met leu lys⁺ ile gly ile cys val gly ile leu gly [leu] ser ile thr ser leu glu ala phe thr gly glu mr leu

α-AMYLASE (B. amyloliquefaciens):

met ile gln lys⁺ arg⁺ [lys⁺ arg⁺] thr val ser phe arg⁺ [leu] val leu met cys thr [leu leu phe] val ser leu pro ile thr lys⁺ thr ser ala↓val

α-AMYLASE (B. subtilis):

met phe ala lys⁺ arg⁺ phe lys⁺ thr ser leu leu pro [leu phe] ala gly [phe] leu [leu leu phe] his⁺ leu val leu ala gly pro ala ala ala mr ala↓

α-AMYLASE (B. licheniformis):

met lys⁺ gln gln [lys⁺ arg⁺] leu tyr ala arg⁺ leu [leu] thr leu [leu phe] ala [leu] ile [phe] leu leu pro his⁺ ser ala ala ala ala↓ala

β-GLUCANASE (B. subtilis):

met pro tyr leu [lys⁺ arg⁺] val leu leu leu leu val thr gly [leu phe] met ser [leu phe] ala val thr ala thr ala↓ser ala↓lys⁺

◄──── HYDROPHILIC SEGMENT ────►│◄──────────── HYDROPHOBIC SEGMENT ────────────►

FIGURE 3: N-terminal amino acid sequences of precursors to extracellular proteins of Bacilli. The amino acid sequences of the N termini of five precursors to extracellular enzymes of Bacilli are shown and compared to the sequence of β-glucanase deduced from the DNA sequence. Published sequences are for β-lactamases of B.licheniformis (8) and B.cereus (7), α-amylases from B.amyloliquefaciens (11), B.subtilis (9,10) and B.licheniformis (31). The arrows show known or postulated cleavage sites. The boxes show some regions of homology which are evident when the sequences are aligned at the boundary of the hydrophilic and hydrophobic regions.

The putative promoter is followed by a sequence (GAAAGGGG) which is complementary to the sequence CUUUCC(U)CC found close to the 3' terminus of the 16S rRNA of B.subtilis. A complex be-

Table I. Codon usage of the β-glucanase gene of B. subtilis

Amino acid	Codon	Number of codons	Amino acid	Codon	Number of codons	Amino acid	Codon	Number of codons	Amino acid	Codon	Number of codons
Phe	UUU	10	Ser	UCU	2	Tyr	UAU	14	Cys	UGU	0
	UUC	4		UCC	2		UAC	3		UGC	2
Leu	UUA	3		UCA	4	Term	UAA	1	Term	UGA	0
	UUG	5		UCG	3		UAG	0	Trp	UGG	8
	CUU	5	Pro	CCU	3	His	CAU	4	Arg	CGU	2
	CUC	1		CCC	0		CAC	0		CGC	1
	CUA	2		CCA	4	Gln	CAA	6		CGA	1
	CUG	2		CCG	2		CAG	1		CGG	1
Ile	AUU	3	Thr	ACU	6	Asn	AAU	8	Ser	AGU	1
	AUC	4		ACC	0		AAC	12		AGC	1
	AUA	0		ACA	12	Lys	AAA	9	Arg	AGA	1
Met	AUG	8		ACG	5		AAG	4		AGG	0
Val	GUU	4	Ala	GCU	6	Asp	GAU	8	Gly	GGU	8
	GUC	5		GCC	3		GAC	5		GGC	4
	GUA	2		GCA	6	Glu	GAA	5		GGA	8
	GUG	2		GCG	1		GAG	2		GGG	4

	NUCLEOTIDES	TTGACA	TATAAT
(a)	435 - 480	AGTTACAGGA<u>TTC</u>A<u>A</u>GTTAGTAAGATTCGATAT	<u>TATC</u>A<u>TT</u>ATTTTG
	468 - 514	TATCATTATT<u>TTGACC</u>GATGTTCCCTTTTGAAAG	A<u>ATC</u>A<u>TG</u>TAAGAT
	485 - 532	ATGTTCCCT<u>TTTGAAAG</u>AATCATGTAAGATCAACA<u>TAG</u>A<u>A</u>AACGCTTT	
(b)	545 - 499	TCCCCTTTC<u>ATTGAAAG</u>CGTTTTCTATGTTGATC	<u>TTACATG</u>ATTCTT
(c)	1334 - 1381	TGAGAAGTTG<u>TTGGC</u>ATGAATCTTCTATTCACT	<u>CATAAT</u>ATCCTATTT

FIGURE 4: Regions showing homology to B.subtilis RNA initiat-
ion sites. On the top line is shown the consensus sequence for
the "-35" and "-10" regions of σ^{55} promoters. Sequences 5' to
the β-glucanase gene are shown in (a) whereas the sequence in
(b) is located on the complementary DNA strand and in (c) down-
stream to the β-glucanase gene. The nucleotides refer to the
numbering given in Fig. 1. Nucleotides that match the consensus
sequence are underlined.

tween these two sequences would have a binding energy $\Delta G = -13.8$

kcals per mole, which is in the range observed for other Bacillus

ribosome binding site sequences (1,25). This sequence is follow-

ed by an ATG ten nucleotides away, which is typical for Bacillus

genes.

There are other sequences to the 5' side of the β-glucanase

gene which resemble elements of σ^{37} promoters but only two such

promoters have been sequenced and the implications of this ob-

servation are correspondingly tentative.

A scan of the strand complementary to the β-glucanase mRNA

sequence revealed a sequence resembling a σ^{55} type promoter

(nucleotides 545 - 499) (Fig. 4) followed by a putative ribosome

binding site, ATG and short (24 codon) open reading frame. The

promoter has -35 and -10 sequences which agree well with the σ^{55}

consensus sequences (Fig. 4) and they are appropriately separ-

ated by 18 base pairs. The ribosome binding site (AAAGGG) is

complementary to the 16S rRNA sequence UUUCCU and is separated

from the AUG by 16 bases. This is longer than the longest

previously reported distance (7). The positions of the β-glu-

canase σ^{55} promoter and this σ^{55} promoter on the opposite strand

would give rise to converging transcriptional complexes. It is

interesting to consider whether these two potential transcript-

ional units may interact and modulate the control of expression

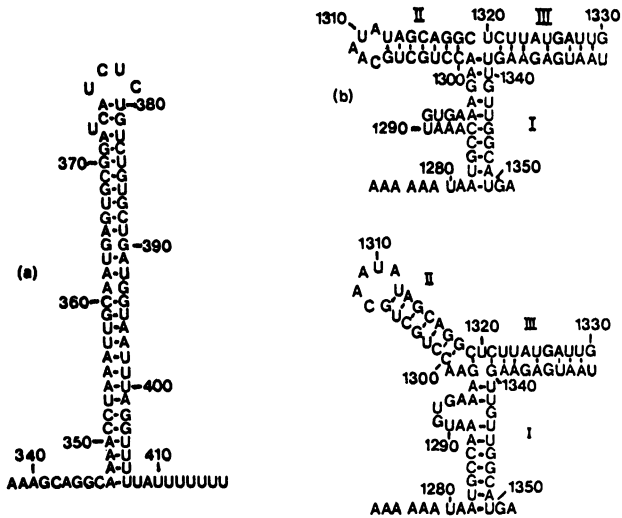of the β-glucanase gene. The 24 codons of the short peptide

FIGURE 5: Schematic drawing of possible mRNA conformations due to inverted repeat sequences. The transcriptional terminator type structure in (a) is 5' to the β-glucanase gene and has a calculated free energy of -18.9 kcal/mol. The attenuator type structure in (b) is 3' to the β-glucanase gene and the two possible conformations have a calculated free energy of -21.4 and -20.4 kcal/mol respectively.

would code for a strongly hydrophobic protein suggesting a role in excretion.

There are other features of this region which suggest that this sequence may not be transcribed. The sequence does not have some of the other elements known to be important in transcriptional initiation sites (1,26). It does not have an AT-rich sequence upstream from the -35 region, and there is no purine 6 or 7 bases downstream of the -10 region, the first purine is 10 bases away. Finally there is no transcriptional terminator-like sequence beyond the short open reading frame.

The sequence was scanned for potential termination signals. There is a relatively long inverted repeat of 28 nucleotides starting at nucleotide 344 (Fig. 5) which has the properties expected of a transcriptional terminator. It lies after an open reading frame (nucleotides 1-255) and before the putative β-glucanase promoter (nucleotides 478-507). The free energy of the stem loop structure shown in Fig. 5 is -18.9 kcal/mol. calculated

according to Tinoco et al. (27) suggesting that this structure
would be stable in solution. It is followed by a sequence of 7
U residues, which is typical of transcriptional terminators in
E.coli (28). It is similar to the sequence of the proposed ter-
minator of the B.licheniformis penicillinase gene (8,29).

The sequence downstream from the β-glucanase gene was exam-
ined for termination signal sequences. There is a complex set of
inverted repeats around nucleotides 1282 to 1350 (Fig. 5) just
after the UAA stop codon of the β-glucanase gene. An inverted
repeat of six base pairs starts at 1300 and a stem loop structure
of these sequences would have a free energy of -21.4 kcal/mol.
A second imperfect repeat of 9 base pairs starts at 1321 but a
stem loop structure of these sequences alone would be unlikely
to form since it would have a positive free energy. Flanking
both these inverted repeats there is a third pair of sequences
starting on the left at nucleotide 1282 and on the right at
nucleotide 1350 which can be arranged in a stem loop structure
with a bulge of 8 nucleotides. These three regions can be arran-
ged in the form of a T structure (Fig. 5). It resembles the
structure proposed for the control region (the translational
attenuator) preceding the MLS resistance gene of plasmid pE194
from S.aureus, which is functional in B.subtilis (30).

There are some interesting features of this proposed
structure.
1). The last nucleotide of the termination codon of β-glucanase
is part of the first stem structure. Ribosomes at this termin-
ation codon would mask a further 12 nucleotides downstream pre-
venting the formation of stem I. Under these conditions stem II
could still form.
2). The bulge loop of stem I (1287 - 1294) contains an AUG
followed immediately by a UGA in phase. A ribosome at the β-
glucanase termination codon UAA might transfer to the AUG which
starts with the last nucleotide of the UAA and then to the AUG
of the bulge sequence, thus de-stabilising the structure of stem
II, and consequently resolving the whole T structure - stem III
is not stable without the contributory forces of stems I and II.
These considerations suggest that this T shaped structure may
have a role as an attenuator modulating the transcription of a

gene downstream from the β-glucanase gene.

The T structure may have a different function as a repress-
or binding site and this is suggested by the observation that
there are sequences in this region at 1344 - 1349 and 1367 - 1372
which resemble the -35 and -10 sequences of B.subtilis σ^{55} pro-
moters respectively (Fig. 4). The -35 type sequence is involved
in the stem I structure in Fig. 5. In the MLS gene of plasmid
pE194 the T shaped structure has been implicated as a binding
site for the 29k MLS protein although it is not known whether
this occurs at the DNA or mRNA level (30).

*To whom correspondence should be addressed

REFERENCES
1.   Moran, Jr., C.P., Lang, N., Le Grice, S.F.J., Lee, G.,
      Stephens, M., Sonenshein, A.L., Pero, J. and Losick, R.
      (1982) Mol. Gen. Genet., 186, 339-346.
2.   Johnson, W.C., Moran Jr., C.P. and Losick, R. (1983) Nature
      302, 800-804.
3.   Lee, G. and Pero, J. (1981) J. Mol. Biol. 152, 247-265.
4.   Pero, J., Lee, G., Moran Jr., C.P., Lang, N. and Losick, R.
      (1982) in Promoters: Structure and Function, Rodriguez, R.
      L. and Chamberlin, M.J. Eds. pp. 283-292. Praeger, N.Y.
5.   Gilman, M.Z., Wiggs, J.L. and Chamberlin, M. (1982) in
      Promoters: Structure and Function, Rodriguez, R.L. and
      Chamberlin, M.J. Eds. pp. 293-304. Praeger, N.Y.
6.   Murray, C.L. and Rabinowitz, J.C. (1982) J. Biol. Chem.
      257, 1053-1062.
7.   Sloma, A. and Gross, M. (1983) Nucleic Acids Res. 11, 4997-
      5004.
8.   Neugebauer, K., Sprengel, R. and Schaller, H. (1981) Nucleic
      Acids Res. 9, 2577-2588.
9.   Yang, M., Galizzi, A. and Henner, D. (1983) Nucleic Acids
      Res. 11, 237-249.
10.   Yamazaki, H., Ohmura, K., Nakayama, A., Takeichi, Y.,
      Otozai, K., Yamasaki, M., Tamura, G. and Yamane, K. (1983)
      J. Bacteriol. 156, 327-337.

11. Takkinen, K., Pettersson, R.F., Kalkkinen, N., Palva, I., Söderlund, H. and Käariäinen, L. (1983) J. Biol. Chem. 258, 1007-1013.
12. Blobel, G. and Dobberstein, B. (1975) J. Cell. Biol. 67, 835-851.
13. Perlman, D. and Halvorson, H.O. (1983) J. Mol. Biol. 167, 391-409.
14. Cantwell, B.A. and McConnell, D.J. (1983) Gene 23, 211-219.
15. Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
16. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
17. Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 2110-2114.
18. Ruther, U., Koenen, M., Otto, K. and Müller-Hill, B. (1981) Nucleic Acids. Res. 9, 4087-4098.
19. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
20. Borriss, R. (1981) Z. Allg. Mikrobiol. 21, 7-17.
21. Davis, B.D. and Tai, P.C. (1980) Nature 283, 433-438.
22. Kreil, G. (1981) Ann. Rev. Biochem. 50, 317-348.
23. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
24. Ikemura, T. and Ozeki, H. (1982) Cold Spring Harbor Symp. Quant. Biol. 47, 1087-1097.
25. McLaughlin, J.R., Murray, C.L. and Rabinowitz, J.C. (1981) J. Biol. Chem. 256, 11283-11291.
26. Le Grice, S.F.J. and Sonenshein, A.B.L. (1982) J. Mol. Biol. 162, 551-564.
27. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.
28. Yanofsky, C. (1981) Nature 289, 751-758.
29. McLaughlin, J.R., Chang, S-Y. and Chang, S. (1982) Nucleic Acids Res. 10, 3905-3919.
30. Horinuchi, S. and Weisblum, B. (1982) in Molecular Cloning and Gene Regulation in Bacilli. Ganesan, A.T. et al. Eds. pp. 287-300. Academic Press Inc. N.Y.
31. Stephens, M., Ortlepp, S.A., Ollington, J.F. and McConnell, D.J. (1984) J. Bacteriol. (in press).