© 1992 Oxford University Press
Nucleic Acids Research, Vol. 20, No. 5 961–974

# Compilation and analysis of DNA sequences associated with apparent streptomycete promoters

William R.Strohl
Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210, USA

Received September 25, 1991; Revised and Accepted January 31, 1992

## ABSTRACT

**The DNA sequences associated with 139 apparent streptomycete transcriptional start sites are compiled and compared. Of these, 29 promoters appeared to belong to a group which are similar to those recognized by eubacterial RNA polymerases containing $\sigma^{70}$-like subunits. The other 110 putative promoter regions contain a wide diversity of sequences; several of these promoters have obvious sequence similarities in the −10 and/or −35 regions. The apparent Shine–Dalgarno regions of 44 streptomycete genes are also examined and compared. These were found to have a wide range of degree of complementarity to the 3′ end of streptomycete 16S rRNA. Eleven streptomycete genes are described and compared in which transcription and translation are proposed to be initiated from the same or nearby nucleotide. An updated consensus sequence for the $E\sigma^{70}$-like promoters is proposed and a potential group of promoter sequences containing guanine-rich −35 regions also is identified.**

## INTRODUCTION

Streptomycetes are mycelial bacteria which produce a wide variety of bioactive compounds including antibiotics, ionophores, immuno-modulators, hydrolytic enzymes, and enzyme inhibitors. Streptomycetes undergo a complex life cycle (1), the initial phase of which is growth as substrate mycelium (2). After growth of the colony on a solid surface has nearly ceased, which may be triggered by a limitation of nutrients, the colonies develop aerial mycelia, utilizing nutrients primarily provided from the hydrolysis of the substrate mycelia (2). The final stage of development is the conversion of the aerial mycelia to spores, which can be dispersed to begin the life cycle over again (1). On solid surfaces, the biosynthesis of antibiotics by streptomycetes typically occurs concomitantly with the development of aerial mycelia and spores (1). Antibiotic biosynthesis and morphological development have been shown to be co-regulated by certain genetic loci (3), whereas other genetic loci regulate either morphogenesis (4,5) or antibiotic biosynthesis (6), but not both. Often the expression of extracellular hydrolases such as proteases also occurs after primary growth has ceased (7). Thus, a vast array of regulatory mechanisms is likely to be required to coordinate the various processes carried out by these organisms. Moreover, the potential for overproduction of antibiotics via genetic manipulation (8,9) and the potential for production of novel products via interspecies cloning (9–12) have brought about a strong interest in how the various genes encoding these processes are regulated. Finally, there is considerable interest in the use of *Streptomyces lividans* as a host for secretion of heterologous recombinant proteins into the culture broth (13–15). Knowledge of promoter structures and transcriptional regulation of genes encoding exported proteins will assist in carrying out these types of experiments.

Detailed analyses of the promoters recognized by RNA polymerase of *Escherichia coli* have shown that two hexamers of nucleotides, centered around −10 and −35 from the transcription start site, are the major recognition sequences for RNA polymerase holoenzymes containing $\sigma^{70}$-like $\sigma$-factors (16–19). Although there are promoter sequences of *E. coli* that are recognized by specific $\sigma$-factors for special purposes, e.g., heat shock response (20), motility (21), and nitrogen metabolism (22,23), there are comparatively few promoters that vary significantly from the $E\sigma^{70}$ consensus sequences (17,18). A recent analysis by Gralla and his associates (23,24) has suggested that the majority of *E. coli* promoters falls into two basic categories: (i) those recognized by $E\sigma^{70}$, the activities of which are modulated by negative and positive regulators that must 'communicate directly' with the RNA polymerase; and (ii) those promoters recognized by $E\sigma^{54}$, which are only regulated by activation, where the location of activator binding is less critical and in some cases may be remote from the binding of the RNA polymerase (23,24). On the other hand, at least eight different RNA polymerase holoenzymes have already been found in *Bacillus subtilis* (16,25), several of which have active roles in the regulation of morphological development of that organism (25,26). For example, four sigma factors of *B. subtilis* are synthesized only during sporulation and a fifth recognizes certain sporulation-specific promoter sequences (25). Similarly, at least seven different RNA polymerase holoenzymes already have been found in the *Streptomyces* (27), suggesting that the sigma factors play an active role in regulation of gene expression in these complex microorganisms. Considerable information is available about *Bacillus* promoter sequences and the corresponding RNA polymerases that recognize them (19,25,28). Relatively limited information, however, is yet available about streptomycete promoters. Reviews by Janssen et al. (29), Hopwood et al. (30), Hutchinson (31), Hütter and Eckhardt (32), and Chater and Hopwood (19), have shown limited comparisons of streptomycete

promoter sequences. Seno and Baltz (33) carried out a detailed analysis on the promoters of 25 streptomycete genes, and many of the analyses made in this paper represent updates on their data. In this paper, I have tabulated and compared 139 streptomycete promoter sequences which are upstream from apparent transcription start sites identified by high resolution S1 nuclease mapping or by primer extension analysis of *in vivo* mRNA. Some streptomycete promoter sequences have been characterized more stringently, i.e., by concurrence of initial *in vivo* data with additional *in vitro* transcriptional analyses (34–42), *in vitro* dinucleotide-primed transcription (36–38), mutational analysis (35, 43–47), DNA footprint analysis using purified RNA polymerase (48), and *in vivo* analysis using promoter-probe vectors (37,49–55).

One note of caution must be included in the treatment of the data included herein. Only a few streptomycete promoters have been studied in enough detail to identify −10 and −35 regions from biochemical and/or genetic evidence. Thus, most of the alignments are made from best comparisons of sequences directly upstream from transcription start sites, rather than on biochemical evidence of their function. Upon more stringent genetic and biochemical analyses, these alignments may prove to be inadequate. It is expected that as additional information based on biochemical and mutational studies of streptomycete promoters becomes available, more definitive conclusions can be drawn concerning the exact structures of streptomycete promoters and structure/function relationships. Nevertheless, this tabulation and analysis confirm some of the generalizations previously made about sequences associated with apparent transcription start sites (29,30,32,33) and suggest some additional sequence relationships among groups of streptomycete promoters. Ultimately, however, promoter classification is based on cognate RNA polymerase holoenzymes, rather than on sequence comparisons such as these. Nevertheless, the purpose of this paper is to provide updated, tabulated information on apparent streptomycete promoters so that future studies on promoters and transcription in streptomycetes have a basis from which to compare new sequences to those already published. Finally, this work is intended to stimulate further efforts in analysis of the various streptomycete promoters and the relationships between promoter structure and gene expression in these unusual differentiating and economically important microorganisms.

## Common features of streptomycete promoter regions

Streptomycete genes have thus far shown a wide diversity in promoter sequences and transcriptional patterns. Of the 87 genes described in this study, 27 of them (ca. 31%) have multiple promoters. Thus far, 13 loci (*vph* [44]; *ermE* [35,43,45]; *aph* [43,45,46]; *kmr* [56]; *redD* [57]; *rep* [40]; *korA/traA* [58]; *sph/orfI* [59]; *xylA/xylB* [60]; *hrdD/bar* [61]; *otrA/otcZ* [62]; *actII(orf1)/actII(orfs2,3)* [63]; *actI/actIII* [64,65]) have been shown to contain overlapping, divergent promoters, a structure which has been postulated to be involved in complex regulatory patterns (46,66). In the cases of *redD* (57), *aph* (46), and *hrdD/bar* (61), the overlapping divergent promoters are located within, or partially within, the open reading frame of the gene. The promoter regions of *mmr* (involved in resistance to methylenomycin) and *orfJI2* (unknown function), are divergent from the same region, but the +1 to ca. −40 regions do not overlap (67). Nevertheless, upstream sequences which may be involved in binding regulatory proteins may overlap in this case (67). Two promoters, *gal-p2* (68) and *nshR-p* (69,70), initiate transcription from the middle of multi-gene operons in which other promoters are found upstream of the first gene in the operon. Both of these promoters apparently provide additional transcriptional capabilities to their respective operons (68–70). Finally, at least six promoters (*orf1590-p2* [55]; *afsB-p* [71,72]; *pIJ101-pc* [58,73]; *brpA-p3* [74]; *aphD-p2* [75]; *hrdD-p1* [61]) lie within open reading frames of streptomycete genes. In one of these cases, the promoter has been postulated to have a specialized function in the temporal regulation of gene expression (55).

Tables 1 and 2 list apparent streptomycete promoter sequences upstream of transcription start sites identified by high resolution S1 nuclease mapping, or primer extension, of *in vivo* transcripts. Transcription start sites analyzed by S1 nuclease mapping procedures are usually indicative of the true transcription start sites; however, since the results of S1 nuclease mapping can also be achieved through post-transcriptional processing or via *in vitro* artefacts (e.g., secondary structure), some caution must be observed in interpreting these data. Moreover, the precise locations of transcription start sites defined by S1 nuclease protection experiments are affected by runs of adenine and thymidine residues in the region of the transcription start site (114). This is probably not a problem with streptomycete sequences, however, because even in the promoter regions streptomycete DNA has a relatively high G+C content. Moreover, multiple adjacent (or nearby) nucleotides are often observed as transcription start sites in S1 nuclease protection studies (115). Thirty-six of 139 streptomycete apparent transcription start sites (26%) were located to multiple adjacent (or nearby) nucleotides rather than to a single nucleotide. The presence of multiple S1 nuclease-generated bands is not uncommon; Brosius et al. (115) suggested that the 5′-terminal triphosphate group of prokaryotic mRNA sterically inhibits S1 nuclease from cleaving the probe precisely at the junction of the DNA/RNA hybrid, often making the observed S1-protected fragment one or more nucleotides larger than the true *in vivo* transcript. Dinucleotide priming of the *dagA* (37) and *ermE* (M.J. Bibb, personal communication) promoters have confirmed that multiple transcription start sites observed *in vivo* occur *in vitro* as well.

The G+C content of DNA isolated from streptomycetes, analyzed by thermal denaturation or buoyant density analyses, is 69 to 78 mol% (116). The coding regions of 27 genes analyzed by Seno and Baltz (33) were found to have an average G+C content of 70.1 mol%. Bibb et al. (35) noted that the G+C content of *E. coli* promoters, measured as 20 nucleotides in both directions from the midpoint between the −10 and −35 hexamers (40 nucleotides total), is 43 mol%, whereas in the *ermE* promoters, the G+C content was found to be 62 to 65 mol%. The apparent streptomycete promoter sequences shown in Tables 1 and 2 have an average G+C content of 57 mol% and 62 mol%, respectively. Moreover, the sequences of 22 promoter regions (from −100 with respect to transcription initiation site to the AUG [or GUG] of the predicted coding regions) were found to average 62 mol% G+C (data not shown). It is not surprising that the promoters shown in Table 1 would have the lowest G+C content of the different analyses, especially since several of these are functional in *E. coli* (see section on $E\sigma^{70}$-like promoters). The promoters shown in Table 2 are, on the average, slightly more G+C-rich than those shown in Table 1. Where tested (discussed later), these promoters are not typically expressed in *E. coli*. Certain of the promoters in Table 2, e.g., *tsr-p2* (47) and *pabS-p* (105), have significantly lower G+C contents than other promoters in that table.

**Table 1.** Apparent streptomycete promoters that are similar to those recognized by *E. coli* $E\sigma^{70}$-like RNA polymerases[a].

| # | Promoter | −35 region −10 region tss[b] | Reference |
|---|---|---|---|
| **SPACER: 16 nucleotides** | | | |
| 1 | SEP3 | TGGGTTGCAGACTCTTGACAACCGCGTAACAGGAGTCATCATATCGCCTATCG | 49 |
| 2 | redD-pr1 | GCAGAGAAGAAGAGTGGTGTAAGCCGTGCACATTGTCATCATGGGCTGCGGGCGC | 57 |
| 3 | vph-pA1 | CACTGGAATGCCCCTACCACGGTTGGTTGTTCGAAACGGGC | 44 |
| **SPACER: 17 nucleotides** | | | |
| 4 | rrnD-p2 | CCCGCAAGAGCCGTTGACACGGAGCGAGCGGGGAGGTAGATTCGAACA | 39 |
| 5 | tra-p (pIJ101B) | CACAGTGTATGCCTTGACACAGCAACTGTGCGACCACAGTATGGACCTTG | 36,58 |
| 6 | mmr-p | CGGCCATCAAAGTTTGACAGCCGTCGTCATATGAGCTTCAGTGAGAACG | 67 |
| 7 | βgal-p | TCCGACGGGGTAATTGATTCGGTTGTGTTCGGGTTCTAGGGTGACCCGT | 76 |
| 8 | korB-p | CAGCCTGAACTAGTTGCGCAGACTGACACAGTCGGTCAGGATGACTTCATG*** | 58 |
| 9 | SEP8 | CAAGCCGCAAGATTGACGAAACATGCAGAGTCCTGCATACTCATGCA | 50 |
| 10 | gyl-p1 | GATAACACAGCTCTTGACGCGCGCGTGACGTCGAACGAGACTCGCGTCCAT | 77 |
| 11 | amlV-p | GGCGTCCGGAGGGTTGACCGGGCGTCGGGGCACTCGTACGGTCACGGCTGAAAACAGTT | 53 |
| 12 | aml-p | CGTGTCCAAAGGGTTGACCGCGGGTACCGCTCGCTCTACGGTCTGCTTCGC | 78 |
| 13 | amySG-p | CGTGTCCAAAGGGTTGACCGCGGGTACCGCTCGCTCTACGGTCTGCTTCGCG | 79 |
| 14 | XP55-p | CTCCCCACCTGGCTTGACGCTTTATTGCGAGTGATGTGCAATAGCTGC | 80 |
| 15 | SEP6 | TCGATGGAATATATGGACAGCGCGTTCATTTCGTGGTTATATGAACGTG | 49 |
| 16 | rrnD-p1 | ACAGCCGCCTGATGTGCATCCACCCCTGCGAGCTGCTAGTGTCCTCTTC | 39 |
| 17 | ssi-p2 | GCATGGGGTCGAGGTGAGTTTCCGCCGGGGACTCGGCAGACTCCGGCACGCCG | 81 |
| **SPACER: 18 nucleotides** | | | |
| 18 | choP-p | GGAACGATCTCGTTGACAGCCTTCACATCGCCTCCATACGGTCATTTC | 82 |
| 19 | dag-p4 | CAGCCGTACCGATTGTCACCCTGCGACACTCCGCTGTAGCATTCGGGAAA | 37 |
| 20 | ermE-p2 | ACGCGGTCGATCTTGACGGCTGGCGAGAGGTGCGGGGAGGATCTGACC | 45 |
| 21 | SEP2 | AGTATTTTTATGTTGACGGTCAAAGTAGCCCTTGAGTAAAATTACCCTCGATA | 49 |
| 22 | ermSF-p | GAGGAAAAGGAGTTGCGGGCGAGCCGTGCCCCTGCTTACCGTTCCCG | 83 |
| 23 | kilB-p1 (pIJ101A) | CGGGGCCTACCATTGCGCAACCTGACACGTGCGTGTCAGACTGATACACG | 36,48,58 |
| 24 | actII(orf1)-p | GGTCCTCGACTATTGGTCCACGAACGACCACCGTTCTACAATGGAACG | 63 |
| 25 | pARC-p | GACACGCTCCCCTTGCCGGTGCTAGCGCGACCGCGCTAGCGTGGTCGG | 84 |
| 26 | cefD-p | ACGACAGGACTCTTGAAGTGCTCTTCGGCTGGTCTTCAGAATCTCTTCGCTATTTT | 85 |
| 27 | mPlc-p | GCTTGTAGATCGTTGACGTGGCACCCTTGTGCACGGCACACTGTCCCT | 38 |
| 28 | aacC7-p1 | CAGAGAAATACGGTGCCGGTGACCGTGAGCGACGGATACCTTCCCGTCCATG*** | 86 |
| 29 | npr-p | TGCATGCCGAATGTGACATGCGCAATCCATGTGGCGTAAAGTCCCGGTG | 87 |

[a]Symbols: The promoters in this table are separated into three sub-groups based on the apparent sizes of the spacer regions between the −10 and −35 regions of 16, 17, or 18 nucleotides. The −35 and −10 hexamers are double-underlined; the transcription start sites are denoted by a dot beneath the nucleotide; and ATG*** indicate promoters in which transcription and translation are proposed to be initiated at the same nucleotide.

[b]Abbreviation: tss, transcription start site.

**Table 2.** Apparent streptomycete promoters, based on S1 nuclease mapping or primer extension analyses, that do not display typical −10 and −35 regions characteristic of promoters recognized by $E\sigma^{70}$-like RNA polymerases[a].

| # | Promoter | −40 * −30 * −20 * −10 * +1 * | Reference |
|---|---|---|---|
| 30 | gal-p1 | TTGTGATGTGACAGGGGGGTGGTGGGTTGTGATGTGTTATGTTTGATTG | 68 |
| 31 | nshR-p | CCGCTCTGGTGGCCGGGGGCGCAGGCTCCCGGCCACTAGACTGCGCGCATG*** | 69 |
| 32 | afsB-p | GCCGAGCGAGCGGGGGGCGGCGCTGTCGCGGCTGCTGGACTTCTACCTG | 71 |
| 33 | gylR-p | CCCGCCGACCATTGGCGGGAGGTCGGCATGGACCGGTAGTGTTCGGCATT | 77 |
| 34 | orfI-p2 | CCTCGGGAGGGAGCGGGTGTTTCTCCCGCAACCTATCGGAGATCGGTACA | 59 |
| 35 | redD-pr4 | GAAGCACCCCACCGGGGTGCGACGGCGGTGGACCGGCATGATGGTGGC | 57 |
| 36 | afsR-p2 | TGCTGACGGTACGGGTGGAGTTGGAGGGGCAAAGCCGAAGAAGCTGATGG | 72 |
| 37 | nshA-p1 | TCGGCCCGGGACGGGGGCGCGCGGGGCCGATCCGTGGCATTGTCGACGAC | 69 |
| 38 | orfI-p1 | CCGTCCGTGGACGGTGGTGGGGCCGAGGACCGCGGGATACGGTCGCCGGCC | 59 |
| 39 | sphP-p | CCGTCCACGGACGGCGGGAGCCTGTACCGATCTCCGATAGGTTGCGCGA | 59 |
| 40 | kgmB-p | CCACGCGGGGACGTCCGAGGGAGCGGGCGAGCCTCGCTAGGCTGGACG | 88 |
| 41 | redD-p11 | GATCACGGCGACCGTGTGCCCCTGCTGCTCCAGGGCCTGGGCGAGCGC | 57 |
| 42 | melC1-p(sta) | CGGTCCGGGCCGATTTCTCCCCTTCTCCTCCGGTCGATAGGTATGCGGGG | 89 |
| 43 | xylA-p | GCTGCTGACATCGGTTTCTCCCTCTTCCGCGGCTCAGCGGGGTGTGTCTCTT | 60 |
| 44 | bar-p1 | TTGTCTCCGTGTCTCCAGCCCCTGTCGGGACTCTCCTCGGGTGCCCTGCAT | 61 |
| 45 | actII(orf2,3)-p | CCACCGTTTCTTGACGCGACCACCGTTCCATTGTAGAACGGTGGTCGT | 63 |
| 46 | rep-p2 | TAAAGCTTTGGTAACGCACCCAGCCTACTCACGTGAGTAGCTTGGAGCG | 40 |
| 47 | tsr-p1 | TGATTGCCGGTCAGGGCAGCCATCCGCCATCGTCGCGTAGGGTGTCACA | 47 |
| 48 | strB-p | GCCGCGGTGGACATATGCCCGAGCGAAGCGGCGCTGCTAGCCTGCGATGA | 75,90 |
| 49 | drrAB-p | GAAGCCATCGCGGCCATGAAGTGTCCTCATTGGGGGCTACGGTACTCAAC | 91 |
| 50 | rep-p1 (cp1) | ACCCAGCCTACTCACGTGAGTAGCTTGGAGCGTGGGCTAGGGTGAGCGA | 40,92 |
| 51 | glnA-p | GAAACAAATGGGTCACGCCCGAGAAATCACCCGTCCCTAGGGTCGAGGAA | 93 |
| 52 | melC1-P(stg) | CGGAGCCAACCGTTCGCGGGAGTGCCCCACGGGCCGGTAGGCATGCGGAG | 89 |
| 53 | aphD-p1 | ACAGCTTTACTTGGCCGTTGCCCGGATGTCCGGGTGCTACTATTCGCGAA | 75,90 |
| 54 | aphD-p2 | CCGTCCGTGCGGCTCACCGAGGACGGTCGGGCGGTGCTACGTTGGCTGAA | 75,90 |
| 55 | pCAT-p | GGCGGAAAATCGCTACGGCCCGCACACCGGCGGCTGATATGCTGAGCCGATG*** | 94 |
| 56 | aacC9-p | GAAAAATTACTCGGTTACCTGACGCCCCGGCTCAGGAGAGCCTGCTAGCTATG*** | 95 |
| 57 | sta-p | CCGGCTGAAACCAGACCTCACCGGGGCAGGCCGGGCATAGCCTCGGGTCATG*** | 96 |
| 58 | korA-p | GTCCATACTGTGGTCGCACAGTTGCTGTGTCAAGGCATACACTGTGCTAG | 58 |
| 59 | orfRP(kmr)-p | TCGAGGCGGTCGAAGACCTCGGCGTACATTTCCGTGA | 56 |
| 60 | ermE-p1 | GCGATGCTGTTGTGGGCTGGACAATCGTGCCGGTTGGTAGGATCCAGCGGTG*** | 35 |
| 61 | gyl-p2 | CGGTCGGCATTGTCGAACACCTACCGGCAATACGCGTTAGAGTGTCCACAGTG | 77 |
| 62 | cp2 | TTTGCCGAAATGGCTCAGCCTACTCACGTGAGTAGTGTAGAGTCGGCTACG | 92 |
| 63 | hrdB-p | CCACGCGGATTGGGCGTAACGCTCTTGGGAACAACACGATGACCTAAGA | 97 |
| 64 | hrpA-p2 | CGGATGACCTTGCCTCGACTGGTTGACTCTCTACACTGAAGATTTACA | 74 |

| # | Promoter | −40 * −30 * −20 * −10 * +1 * | Reference |
|---|---|---|---|
| 65 | tipA-p | CGTCCGGGCTTGCACCTCACGTCACGTGAGGAGGCAGCGTGGACGGCG | 98 |
| 66 | est-p | TCACCTCCATTGGATTATCACGCTCCGTATGTGACTGAATGCAGAGCG | 99 |
| 67 | tlrC-p | GCACCGGGCCTTGGTGATCACGTACGTTGACCTTCATGGGAACACTG | 100 |
| 68 | brpA-p3 | GACGGACCGGCGTCGCCGGCGCAGCATTTGCCTCCGGACTTCGAGCGGCAT | 74 |
| 69 | amy-p2 | TCGCAACTCACCGAAATGTCTTGCAGAAAGCCTTGACGCTGTTAACGACA | 101 |
| 70 | orf1590-p2 | GGGTCAACGACCTCGGAGCACAGCGCGGTCTGACACTTCGCTACGACAAG | 55 |
| 71 | orf-p1 | ACCGCGTGCACCTGCGATCGCCGATCAACCGCGACTAGCATCGGGCGCA | 35 |
| 72 | kamB-p | ACGACGTGGTGCTCGACGTCGGCACCGGCGACGGGAAGCATCCGTACAA | 102 |
| 73 | tylF-p | CGGGCTCCTCGTTCCGGCGC....GCGCCGATAGCGTCCGTCCTC | 33 |
| 74 | saf-p | GGAAACGGTGGTCCGTTTCCCGCCCCTGCCCGTAGGCCGTGCGCGTCCCGC | 103 |
| 75 | actI-p | CGTCACCAGTGCGACTTCGGAGTCCTGCGTGGCCATGTGTTCCCCTCCCT | 64 |
| 76 | casA-p | AGCGTGCGCGAGGGCTTCACCGCATGGCGCCCCGCCGTTACCGTGCCCT | 104 |
| 77 | gal-p2 | CCTGGAACTTTTCACTTCCGCCGTACGTCCGGCAAGCTGAAGTTCCTCG | 68 |
| 78 | cp3 | AGCTGTGCGCCGGGTGCTCGAAGGTCGTGCCGACCGGTGAGGTTTCCGTCTC | 92 |
| 79 | cp3′ | TGCCGACCGGTGAGGTTTCCGTCTCCGAAGAGTCCGTTGAGGTTCCGGGC | 92 |
| 80 | cp3" | AAGAGTCCGTTGAGGTTCCGGGCCTGAGTATGACGGTCAGTCAGAAGAGC | 92 |
| 81 | cp3"′ | GGGCCTGAGTATGACGGTCAGTCAGAAGAGCTACACGCCCGTTGAGGGC | 92 |
| 82 | pA1 | TCCGCGCCTTTCGTCGGGGCCCGTAGGGGTTTCGGACATTCTTGTGCGGGG | 46 |
| 83 | afsR-p1 | GCGTCTCCCACGGCTGACGTGGTCGGCATGAACAAGGCAAACTGACGTG | 72 |
| 84 | amy-p3 | CCTCTTTCGCAACTCACCGAAATGTCTTGCAGAAAGCCTTGACGCTGTT | 101 |
| 85 | dag-p2 | CACGTGGGCGTTCCGGAACTTTTTGCACGCACGCGAGCTCTCGAATTTT | 37 |
| 86 | actIII-p | CCCATCTCCCTTCGACCGCCGCTCGAGCCGGGCGCCCAAAGCTGAGCGCG | 64,65 |
| 87 | rep-pA2 | GTGCTTCGCTCACCCTAGCCCACGCTCCAAGCTACTCACGTGAGTAGGC | 40 |
| 88 | orf-p2 | CGGTCAGATCCTCCCCGCACCTCTCGCCAGCCGTCAAGATCGACCGCG | 45 |
| 89 | pabS-p | GGGGCATGGTACCCCCACATCTATTGAATCCGCAACGCGCAGTATCATG | 105 |
| 90 | bar-p2 | ATTCCCAACCTGCCCCTCGATTTTTCTGATCATGCAGTACCCTGTGCCGCCAC | 61 |
| 91 | xylB-p2 | GTAGTTCATGCCGCGGCTCCTTGCTCGCTGAGGCTATTTCGTCATGGCC | 60 |
| 92 | orfJI2-p | AACTTTGATGGCCGTCAACATTTGATGGCTGTCGTACCATGGTGGAAC | 67 |
| 93 | aph-p1 | GACGAAAGGCGCGGAACGGCGTCTCCGCCTCTGCCATGATGCCGCCCATG*** | 46 |
| 94 | rph-p | GTCAAATCACCTAGGGAGAAGGTGCCTTCTCTGCCATGATGCCGACCATG*** | 106 |
| 95 | afsA-p | GGGGAGTTATGCCCGAAGCAGCAGTCTTGATCGATCCGGTGCCGACTATG*** | 107 |
| 96 | vph-p2 | GGGAGCGACGGAATGCGCGTGCACCGCCCGCTGCCCGTTTC | 44 |
| 97 | cdh-p | GACAATCGGCCTCGAAACTGGAACCTGTTTCAGTTAAGCTGCCCGTCATG*** | 108 |
| 98 | hyg-p | GAACGTCCCCGACGTGGCCGACCAGCCCGTCATCGTCAACGCCTGACCG | 109,110 |
| 99 | aacC7-p2 | CCCGTGCGGCCCTGTCAGTTGCGCCTCCGCGATCGGTTACGCCCTGACC | 86 |
| 100 | pac-p | CTCGCGCCCCCGCGGCCGCCAGCACCGGTGCCGCCACCATCCCCTGACCC | 111 |

| # | Promoter | −40 * −30 * −20 * −10 * +1 * | Reference |
|---|---|---|---|
| 101 | kmr-p | CTCGGCGCCCACTCCTCCACCAGGGAGCGGCTCCCCATC | 56 |
| 102 | redD-pr2 | AAGAGGAAGATGAACGGGATGTTCACCGCGGTCAGAGGGCAGAGAAGAAG | 57 |
| 103 | endoH-p | ATTGACTGATTGACGCGCTTCCGGCGGGCAGGGGAGGCACGG | 34 |
| 104 | vph-p1 | GCAGCGCCGTGTGCGGCCTGCCCCGGCCCGGCGGGAGCGACGGAATG | 44 |
| 105 | pA2 | GGTACTTCCGGCGCAACGTGCTGTCGTCCATGGGCGGCATCATGGCAGA | 46 |
| 106 | rrnD-p3 | AAACGAAGGCCGGTAAGACCGGCTCGAAAGTTCTGATAAAGTCGGAGCC | 39 |
| 107 | rrnD-p4 | GGAAAGCGCCGAGGAAATCGGATCGGAAAGATCTGATAGAGTCGGAAAC | 39 |
| 108 | orf1590-p1 | GGAGAACTACGCTGTGTTTACTGGTGTTCTCGACAGGGGGGCATATTCCT | 55 |
| 109 | A1-p | TGCAGGCGGGGAACTCCTATGCCGACCTCACGAGCCACGGCGGCAGCGG | 42 |
| 110 | pIJ101-pc | AGCTGGACGAGATTGAGAAGGAGGCTGCGCCCGCTGATCGCGGAGGCGG | 73 |
| 111 | kilB-p2 | CGACTCAGCACCGCCGCATGACGGATGGCCCCGCCGGAAACCGGCGGGGC | 58 |
| 112 | mPtk-p | ACTGTCCCTGCCATGTGTGACCTCGGTCCCCCGCGGGCACACAGGGAAGGC | 38 |
| 113 | bla-p | GCGCACTCGGGTCTTCGACCGTATGTCCGCACCGGGACGAGAGTGTTCTCG | 112 |
| 114 | aph-p2 | GCGCGGTGGGGATTCCGGCCGAACGCGCCGACGCCCATGTGACCGCCT | 46 |
| 115 | dag-p3 | TACCTCCTGGAGCCTAGCTCCTCCTGCGCCGTGGAATGATCGTGCCAC | 37 |
| 116 | otrA-p | GTCGCCGTCAAGCCCTGACCCTGCGTGGCCGCCCTTGCTACCGTGATCA | 62 |
| 117 | pA4 | GTTCACTGCGTGCCACTCGTGGTGCGGGTACTTCCGGCGCAACGTGCTG | 46 |
| 118 | rep-pA1 | GTGAGTAGGCTGGGTGCGTTACCAAAGCTTTACCTTCGGAACACAAGAA | 40 |
| 119 | dag-p1 | TTTGGCGCCCAGGGTCTGCGGAAGTCATTGCCAAATATAAGATTCTTCA | 37 |
| 120 | orf-p3 | ATCCTGGTTGCGGCGACGCGGGCGCGGCTGCTCGTCCGAACTGCTCACC | 45 |
| 121 | redD-pr3 | GGTGGGGGCCGCAGATGTTTGCCCGTCGAGCCGAAAGAGGAAGATGAAC | 57 |
| 122 | tsr-p2 | TTGGCTCGACGCAGCCCAGAAATGTATGATCAAGGCGAATACTTCATAT | 47 |
| 123 | xylB-p1 | AAATTAGTATGCAGGAGCACTTCTGGGAAGAGACACACCCCGCTGAGC | 60 |
| 124 | nshA-p2 | CGGCTCCGGTGCAGGCATGATCCGCAAAAGGGTGGAAAGTCGAATTAC | 69 |
| 125 | sapA-p | ACATCTGCCAACGACGTAACAACCCCCGAAGGTGCAAAGGTCTCAACTGGT | 52 |
| 126 | dac-p | TCCTTCGCGTGACATGCAACCCATCTGCCCCTCCTGCGCGTAGAGATGGTGCC | 113 |
| 127 | amy-p1 | TCTTGCAGAAAGCCTTGACGCTGTTAACGACATGAACGGCAGGCTCCGGTA | 101 |
| 128 | hrdD-p2 | CGAGGGGCAGGTTGGGAATTCTGTCCGGATTCCAGTCGTTGTTTCCAT | 61,97 |
| 129 | brpA-p1 | CCTTGCCTCGACTGGTTGACTCTCTACACTGAAGATTTACATCTGATT | 74 |
| 130 | pA3 | GCGTGCCACTCGTGGTGCGGGTACTTCCGGCGCAACGTGCTGTCGTCCA | 46 |
| 131 | hrdD-p1 | ACTCCGCGTCCGTGGCAACCCTCAGGCGGTACGGGCCGTCTTCAGGGT | 61,97 |
| 132 | ssi-p1 | TCAGGCCGATTAAGAGGCGGCGGATATTCGGCCATCTGGCCACTTCGCT | 81 |
| 133 | pA5 | CCGGTGAGCTGGTAGACGAAGGCGCCCCGAGTCTCCTTCGTTCACTGCGT | 46 |
| 134 | mP2/2′-p | CCCGCGGGCACACAGGGAAGGCAGCGCCGGGACAAGGTGTTGCACGATAGGTG | 38 |
| 135 | mP1a-p | CCATCGAGAACTGGCCCTTGGCGTCCAGCGGCTTGTAGATCGTTGACGT | 38 |
| 136 | mP1b-p | GCCCTTGGCGTCCAGCGGCTTGTAGATCGTTGACGTGGCACCCTTGTG | 38 |

| # Promoter | −40 * −30 * −20 * −10 * +1 * | Reference |
|---|---|---|
| 137 otcD-p1 | CGCTGTGGGATCCGCCCGGCGGGAACACACCCGCGGGCATGATCACGGT | 62 |
| 138 otcD-p4,5,6 | CGCATTTCGTTCCGCCCTCTCTGACGTCGCTGTGGGATCCGCCCGGCGG | 62 |
| 139 otcD-p2,3 | CCGCCCTCTCTGACGTCGCTGTGGGATCCGCCCGGCGGGAACACACCCGCGGGC | 62 |

[a]Symbols: The relative distances from the transcription start sites at +1 (and denoted by a dot beneath the nucleotide) are given for easy comparison. Sequences of similarity and the possible −10 regions of promoters #29−37 are underlined; ATG*** or GTG*** indicates the promoters in which transcription is initiated at the same nucleotide.

### Relationship of transcription start site and translation initiation codon

Of 48 streptomycete genes analyzed, the distance from the transcription start site to the coding region (not including the eleven genes containing leaderless mRNAs [described below]) ranged from 9 to 345 nucleotides. With most of the streptomycete genes analyzed, the transcription start sites were within 100 nucleotides of the apparent coding regions. This compares to an average distance of ca. 23 nucleotides separating the transcription start site from the coding region in a typical *E. coli* gene (117). For three genes, the nearest detectable transcription start sites were 298, 335, and 345 nucleotides upstream of the coding regions. In a few of these genes with extraordinarily long 5′ untranslated sequences, the sequences between the transcription start sites and coding regions had significant secondary structures that were hypothesized to contain regulatory sites such as antiterminators (69,118).

The apparent Shine−Dalgarno (119) sequences of 44 streptomycete genes, chosen to reflect genes encoding a wide range of functions, were analyzed for their complementarity to the 3′ end of the 16S rRNA of *S. lividans* (120) and for their distance upstream of the initiation codon, as shown in Table 3. These Shine−Dalgarno sequences ranged from 5 to 12 nucleotides (average of 8.5 nucleotides) upstream of the initiation codons. *E. coli* Shine−Dalgarno sequences are typically 5 to 9 nucleotides upstream of the initiation codon (125), and *B. subtilis* Shine−Dalgarno sequences are 7 to 14 nucleotides upstream of the AUG (122). McLaughlin et al. (126) proposed that the Shine−Dalgarno sequences of Gram-positive bacterial mRNAs are typically able to form strong complexes with the 3′ end of 16S rRNA, whereas *E. coli* Shine−Dalgarno sequences are more variable. As can be seen in Table 3, the strength of the measured apparent streptomycete Shine−Dalgarno sequences varied considerably, from a low of −2.2 Kcal/mol to a high of −22.2 Kcal/mol (using Tinoco's rules [123] for comparison to other published values; average, −11.3 ± 5.1 Kcal/mol [± 1 SD]; n = 44]), as has been observed also with Shine−Dalgarno sequences in *E. coli* (127). The conserved Shine−Dalgarno sequence for these 44 streptomycete genes was (a/g)-G-G-A-G-G. Previously, studies have found that in *E. coli* a rather poor Shine−Dalgarno region is tolerated (two Gs and an A, yielding −7.2 Kcal in binding energy, will suffice; 127) but in *B. subtilis* a rather extensive complementarity between the 3′ end of the 16S rRNA and the Shine−Dalgarno region is required (e.g., −14 to −23 Kcal/mol) for translational initiation (122). The ability of streptomycetes to express *E. coli* genes, such as the *ampC* gene, which has a relatively poor Shine−Dalgarno sequence (TATGGAA [128]), further suggests that streptomycete ribosomes do not require extensive complementarity between the Shine−Dalgarno sequences and the 3′ end of the 16S rRNA in order to initiate translation. Expression of foreign genes containing poor Shine−Dalgarno sequences and the presence of a wide range of Shine−Dalgarno sequences in streptomycete genes (Table 3) indicate that the streptomycete genes that have been analyzed to date do not fit the rules proposed by McLaughlin et al. (126) for the structure of Shine−Dalgarno sequence regions of Gram-positive bacteria.

In the case of eleven actinomycete genes, i.e., streptothricin acetyltransferase (*sta*) from *S. lavendulae* (96), the erythromycin resistance gene (*ermE*) from *Saccharopolyspora erythraea* (formerly *Streptomyces erythreus*) (35,43,45), the aminoglycoside phosphotransferase gene (*aph*) from *S. fradiae* (43,45,46), ribostamycin phosphotransferase (*rph*) from *S. ribosidificus* (106), aminocyclitol acetyltransferase (*aacC7*) from *S. rimosus* forma *paromomycinus* (86), the nosiheptide resistance (*nshR*) gene of *S. actuosus* (69,70), *afsA* from *S. griseus* (107), chloramphenicol acetyltransferase from *S. acrimycini* (94), *korB* from plasmid pIJ101 (58), as well as *aacC9* from *Micromonospora chalcea* (95) and *cdh* from *Norcardia* spp. (108), transcription from the nearest promoter and translation are proposed to be initiated at the same nucleotide (or in one case one nucleotide removed from the promoter; 35) (Table 4). In the case of the *cdh* gene from *Nocardia*, transcription was initiated from the same nucleotide in both *Nocardia* and *S. lividans* (108). The transcription start sites of these eleven genes were characterized using S1 nuclease protection experiments and, in some cases, *in vivo* promoter probe analyses. *In vitro* transcription analyses and mutational studies of *aph-p1* (43,45,46) and *ermE-p1* (35,43,45) confirmed that those transcripts were not the result of post-transcriptional mRNA processing. Edman degradation of the purified *rph* (106), *sta* (96), *aph* (129) and *cdh* (108) gene products also confirmed the N−terminal amino acid sequences as those which were expected from the nucleotide sequences. The unusual structures of these eleven streptomycete genes are especially interesting since eight of them are involved in secondary metabolism, i.e., antibiotic resistance or differentiation (as in the case of *afsA* [107]). Several of these genes do not possess obvious Shine−Dalgarno sequences upstream of the translation initiation site or within the first ca. 20 nucleotides of the coding regions. The *aph* (46) and *rph* (106) genes are strongly expressed in *S. lividans*, indicating that *S. lividans*, and presumably most streptomycetes, can translate mRNAs with this structure. This gene structure displays a relationship between transcription and translation which is strikingly different from the common promoter-ribosome binding site-translation initiation sequence of most prokaryotic genes

**Table 3.** Analysis of possible Shine–Dalgarno regions in 44 streptomycete genes[a].

| # | Gene | Shine–Dalgarno site and initiation codon | Distance (nt) from initiation codon | Binding strength (ΔG in Kcal/mol) | Reference |
|---|---|---|---|---|---|
| 1. | dag | AAGAAGGAGAACGAUCGUG | 11 | −12.8 | 37 |
| 2. | rep | AAGGGGCGGGAACAUG | 8 | − 8.4 | 40 |
| 3. | XP55 | GUGGGGGAGACAUG | 6 | −12.2 | 80 |
| 4. | amlV | CAGGAGGAAUCAUG | 6 | −16.6 | 53 |
| 5. | aml | CAGGAGGCACCACAUG | 8 | −16.6 | 78 |
| 6. | amySG | CAGGAGGCACCACAUG | 8 | −16.6 | 79 |
| 7. | ssi | CGGAAGGAUGCACACAAUG | 11 | − 7.2 | 81 |
| 8. | cho | UGAAAGGGCAUACAUG | 8 | − 9.0 | 82 |
| 9. | ermSF | UGAGAGGUGGUCCUCAGUG | 11 | −16.0 | 83 |
| 10. | amy | GACGAAGGAGCCACAAGAUG | 12 | − 2.2 | 101 |
| 11. | gylR | ACGGAGGCAGUACGUCGAUG | 12 | −14.4 | 77 |
| 12. | est | UGAAAGGGCACAGCCAUG | 10 | − 7.2 | 99 |
| 13. | korA | UCGAAGGAGUCGUCAUG | 9 | − 7.2 | 58 |
| 14. | aphD | UUGAAGGGUGUGUAAUG | 9 | − 7.2 | 75 |
| 15. | galE | CGAGAGGUAGCGAGUUCAUG | 12 | −11.6 | 68 |
| 16. | gyl | AAGGAGUCGCGGGUG | 7 | −14.0 | 77 |
| 17. | cefD | CGGGAGAUGCGUUGACAUG | 11 | −11.6 | 85 |
| 18. | glnA | UAGGAGGAGCUGGAUG | 8 | −16.6 | 93 |
| 19. | orfI | AAGGAGUUGAUCGAUG | 8 | −22.2 | 59 |
| 20. | bla | CAGGAGGUCCGGACAUG | 9 | −18.8 | 112 |
| 21. | galU | GAGGAGUGCGGCAGUG | 8 | −11.6 | 68 |
| 22. | afsR | AGGGGGACGGCAUG | 6 | − 5.0 | 72 |
| 23. | orf1590 | CGAGGGGUGGCGCAUG | 8 | − 9.4 | 55 |
| 24. | hyg | AUAGAGGUCCGCUGUG | 8 | −15.0 | 109 |
| 25. | actIII | AGGGAGGGGAACACAUG | 9 | −16.6 | 65 |
| 26. | dac | CGGGAGAAGAAUCAGAUG | 10 | −11.6 | 113 |
| 27. | sapA | AUCGAGGUGCCAUG | 6 | −13.8 | 52 |
| 28. | tsr | CCGGUAGGACGACCAUG | 9 | − 5.0 | 121 |
| 29. | pAC | AAGGAGACCUUCCAUG | 8 | −15.0 | 111 |
| 30. | sph | CCCGAGGAAUUCGAUAUG | 10 | −11.6 | 59 |
| 31. | nshA | GAGGAGGAGGACCCGUG | 9 | −16.6 | 69 |
| 32. | pIJ101A (kilB) | CAGGGGGCUCACAUG | 7 | −12.2 | 58 |
| 33. | tra | CUCGACGACCAUG | 6 | − 2.2 | 58 |
| 34. | aacC7[b] | CGCGACGCUGAUG | 6 | − 2.2 | 86 |
| 35. | drrAB | CUGGGGGCGUUAGGUG | 9 | −10.0 | 91 |

| # | Gene | Shine–Dalgarno site and initiation codon | Distance (nt) from initiation codon | Binding strength (ΔG in Kcal/mol) | Reference |
|---|---|---|---|---|---|
| 36. | brpA | GAGGGGGCCGUG | 5 | -12.2 | 74 |
| 37. | strB | AUGGAGGAGAGUCAUG | 9 | -16.6 | 75 |
| 38. | βgal | CGGAAGGCCACGGUCAUG | 11 | - 7.2 | 76 |
| 39. | orfP(erm) | ACGGACACUCGCAUG | 8 | - 7.2 | 35 |
| 40. | npr | CCGCAGAAAGCAUG | 7 | - 2.2 | 87 |
| 41. | melC1(sta) | CAGGAGGUCCCGCAUG | 9 | -18.8 | 89 |
| 42. | melC1(stg) | CCGGAGGUCCGUAUG | 8 | -16.6 | 89 |
| 43. | kamB | CAAGAGCCGGAUG | 6 | - 4.4 | 102 |
| 44. | tlrC | CAGGGGCUUUCGCAUG | 9 | - 7.2 | 100 |
| Conserved: | | .....GGAGG..... | Ave.: 8.5±1.8 nt | Ave.: -11.3±5.1 | |
| 16S rRNA[c]: | | 3'-OH-UCUUUCCUCCACUAG-5' | Range: 5 - 12 nt | Range: -2.2 to -22.2 | |

[a]The underlined nucleotides indicate complementarity to the 3' end of the 16S rRNA from *S. lividans* (120). The distance of the Shine–Dalgarno region from the initiation codon was measured as the first base to the right of the AGGA or its equivalent (122). The ΔG of binding was calculated using Tinoco's rules (123) rather than the updated rules of Turner et al. (124) for direct comparison to previously published ΔG values for Shine–Dalgarno sequences of *E. coli* and *B. subtilis* (see text). The initiation codons are designated as AUG or GUG.
[b]This represents the Shine–Dalgarno region upstream of the gene preceding the *aacC7* resistance gene in *S. rimosus* forma *paromomycinus*.
[c]The 3' end of the 16S rRNA sequence of *S. lividans* was determined by Bibb and Cohen (121).

(127). Other prokaryotic genes in which transcription and translation are initiated from the same nucleotides include the lambda *cI* gene transcribed from the *prm* promoter (130), the *tetR* gene in transposon *Tn1721* (131), and the bacteriorhodopsin (*bop* [132]), halo-opsin (*hop* [133]), bacteriorhodopsin-related (*brp* [134]) genes in *Halobacterium halobium*, and the *polA* gene of *Streptococcus pneumoniae* (135). Significantly, the *polA* gene was shown to be transcribed in *S. pneumoniae*, *B. subtilis*, and *E. coli* using the same transcription start site (135).

### Streptomycete promoters with sequences similar to *E. coli* $E\sigma^{70}$-like promoters

Of the 139 apparent streptomycete promoter sequences tabulated, 29 appear to fall into a group which has been previously described as streptomycete *E. coli*–like promoter sequences (30). These promoters typically are associated with 'housekeeping' genes (19), but as can be seen in Table 1, a wide variety of genes associated with secondary metabolic functions is also included in this group. Included in this group are the 'SEP' promoters, which have been defined functionally only by their activity in *E. coli* (49,50); it is possible that these 'SEP' promoters may not have activity *in vivo* in streptomycetes.

The promoters in Table 1 were grouped together based on two criteria: (i) their sequences are relatively similar to the consensus sequence for *E. coli* promoters recognized by $E\sigma^{70}$ (in which the $-35$ hexamer is TTGACA and the $-10$ hexamer is TATAAT [17,18]); and (ii) the $-10$ and $-35$ hexamers of these promoters are 16 to 18 nucleotides apart. The latter requirement is especially important because >92% of all *E. coli* promoters recognized by $E\sigma^{70}$ have spacers between $-10$ and $-35$ of 16 to 18 nucleotides (57% of those have spacers of 17 nucleotides [18]), spacer mutations have been shown to yield strong effects on promoter activities (136,137), and the RNA polymerase holoenzyme appears to contact only one side of the DNA helix (138). Recent mutational studies on sigma factors (139–141) have confirmed the premise of Losick and Pero (26) that sigma factors generally contact nucleotides in both the $-10$ and $-35$ hexamers.

The $E\sigma^{70}$-like streptomycete promoters have been separated into three subgroups on the basis of the number of nucleotides in the spacer regions separating the $-10$ and $-35$ hexamers. If promoters with a 19 nucleotide spacer region were included, certain other streptomycete promoters such as *amy-p1*, *hrdD-p1*, and *gylR-p* might be included with the other apparent $E\sigma^{70}$-like promoters. The average distance between the $-10$ and $-35$ hexameric regions of the promoters in Table 1 is 17.3 nucleotides. This is similar to *E. coli* promoters recognized by $E\sigma^{70}$, in which the optimal spacer distance is $17 \pm 1$ nucleotides (17,18). Jaurin and Cohen (128) found that addition of 1 nucleotide to the 16 nucleotide spacer region of the *E. coli ampC* promoter increased transcription efficiency 16-fold in *E. coli* and 30-fold in *S. lividans*. Thus, it is apparent that the size of the spacer also is important to the activity in *S. lividans* of $E\sigma^{70}$-like promoters.

Certain of the putative promoters shown in Table 1, including *SEP2* (49), *SEP3* (49), *SEP6* (49), *SEP8* (50), *pIJ101A-p* (36) (*kilB-p* [58]), *pIJ101B-p* (36) (*tra-p* [58]), *pIJ101-pc* (73), and *XP55-p* (80), are functional in *E. coli*, further demonstrating the relationship between these and *E. coli* promoters recognized by $E\sigma^{70}$. S1 nuclease protection analyses showed that the transcription start sites of *pIJ101-pc* (73), *ampC* (an *E. coli* gene [128]), and *pIJ101A-p* (36) were at the same nucleotide or within

**Table 4.** Promoter regions and 5′ coding regions of streptomycete genes in which transcription from the nearest promoter and translation are proposed to initiation from the same nucleotide[a].

| # | Gene | −35 region    −10 region | tss/TIC[b] | Coding region | Reference |
|---|---|---|---|---|---|
| 1. | ermE | GATGCTGTTGTGGGCTGGACAATCGTGCCGGTTGGTAGGATCCAGCG | GUG | AGC UCG GAC GAG CAG CCG CGC | 35 |
| 2. | korB | AGCCTGAACTAGTTGCGCAGACTGACACAGTCGGTCAGGATGACTTC | AUG | ACG CAA AAG ACA CCG GGC GAG | 58 |
| 3. | nshR | CGCTCTGGTGGCCGGGGGCGCAGGCTCCCGGCCACTAGACTGCGCGC | AUG | ACU GAG CCC GCC AUC AUC ACG | 69 |
| 4. | sta | GGCTGAAACCAGACCTCACCGGGGCAGGCCGGGCATAGCCTCGGGTC | AUG | ACC ACG ACC CAU GGC AGC ACG | 96 |
| 5. | pCAT | CGGAAAATCGCTACGGCCCGCACACCGGCGGCTGATATGCTGAGCCG | AUG | GAC GCC CCG ATC CCG ACC CCG | 94 |
| 6. | aacC9 | AAATTACTCGGTTACCTGACGCCCCGGCTCAGGAGAGCCTGCTAGCT | AUG | GAA GAG AUG AGC UUA CUC AAU | 95 |
| 7. | afsA | GGGGAGTTATGCCCGAAGCAGCAGTCTTGATCGATCCGGTGCCGACT | AUG | GAC GCG GAG GCC GAG GUG GUG | 107 |
| 8. | aacC7 | GAGAAATACGGTGCCGGTGACCGTGAGCGACGGATACCTTCCCGTCC | AUG | GAC GAA CUC GCC UUG CUC AAG | 86 |
| 9. | aph | GACGAAAGGCGCGGAACGGCGTCTCCGCCTCTGCCATGATGCCGCCC | AUG | GAC GAC AGC ACG UUG CGC CGG | 46 |
| 10. | rph | GTCAAATCACCTAGGGAGAAGGTGCCTTCTCTGCCATGATGCCGACC | AUG | GAA AGC ACG UUG CGC CGG ACA | 106 |
| 11. | cdh | GACAATCGGCCTCGAAACTGGAACCTGTTTCAGTTAAGCTGCCCGTC | AUG | GGU GAC GCA UCU UUG ACC ACC | 108 |

[a]Symbols and abbreviations: The transcription start sites are denoted by a dot beneath the nucleotide, and GUG and AUG indicate the initiation codons.
[b]Abbreviations: tss, transcription start site; TIC, translation initiation codon.

1−2 nucleotides in *S. lividans* and *E. coli*, indicating that the transcription initiation sites for these genes in these two very different organisms were nearly identical. Interestingly, the G+C content of these putative promoters (*SEP2*, *SEP3*, *SEP6*, *SEP8*, *pIJ101A-p*, *pIJ101B-p*, and *pIJ101-pc*, and *XP55-p*) transcribed in *E. coli* is 47.9 mol%. On the other hand, the *vph*, *tsr*, *aph*, and *ermE* genes were not expressed from their own promoters (shown in Table 2) in *E. coli* (32,35). As would be expected, most of the promoters (e.g., *ermE-p1*, *aph-p1*, *aph-p2*, *tsr-p1*, *tsr-p2*, *vph-p1*, *vph-p2*) of these genes do not have sequences resembling E$\sigma^{70}$-like promoters, so their function in *E. coli* would not be expected. Why *ermE-p2* (Table 1; spacer, 18 nucleotides) is not active in *E. coli* is not known at this time.

A compilation of −10 and −35 hexamer sequences of the streptomycete E$\sigma^{70}$-like promoters indicate that the updated consensus sequences for the these now should be T-T-G-A-C-(Pu) for the −35 region and T-A-g-(Pu)-(Pu)-T for the −10 region (Table 5).

## Other streptomycete promoters

Only 29 out of the 139 streptomycete promoters analyzed appeared to fit into an E$\sigma^{70}$-like promoter consensus group. An effort has been made to group together promoters with similar sequences. In obvious cases, the similar sequences between two promoters, or among several promoters (particularly in the important −10 and −35 regions), are indicated by underlining. These side-by-side comparisons, however, are not intended to depict classes of promoters. Indeed, such classes can only be constructed after it has been determined which RNA polymerase holoenzymes recognize and transcribe which sequences. Nevertheless, among the remaining 110 promoters, several interesting comparisons can be made which may suggest possible relationships for future studies on their functionality. For example, *gal-p1*, which is now known to contain a string of G residues in the functional −35 region (Westpheling, pers. comm.) has a similar −35 region with several other promoters, including *nshR-p*, *afsB-p*, *gylR-p*, *orf1-p1/p2*, *sph-p*, *afsR-p2*, and *nshA-p*. In several of these promoters, a possible E$\sigma^{70}$-like −10 region can be seen (as marked), although the functional significance of this is as yet unknown. Other apparent streptomycete promoters having similar sequences can be compared directly by viewing Table 2.

## Promoters with specialized structures or functions

Considering the broad ranges of metabolism that are encompassed by the promoters described, it is likely that sequences other than just the −10 and −35 regions are important to the regulation of transcription. For example, promoters *sapA-p* (52) and *tipA-p* (98) are not constitutively transcribed; *sapA-p* is temporally regulated and is maximally transcribed after exponential phase is completed (51,52) and *tipA-p* is induced by thiostrepton (98). Recent studies have also shown that activity of *aphD-p1* (the promoter upstream of *strR*) is dependent on the presence of A-factor (90). Although specialized RNA polymerases may recognize features of the −10 and/or −35 regions of these genes, other features of the promoter regions also may be involved in temporal or induction/repression mechanisms. DNA within several streptomycete promoter regions has significant potential for forming secondary structures (i.e., hairpin loops), including *tipA-p* (98) mentioned above, *pabS-p*, which is regulated by phosphate (105), *hyg-p* (110), and *rep-p1*, the promoter for the repressor gene of $\phi$C31 (40).

Some streptomycete promoters thus far studied are clearly regulated temporally. *In vivo* promoter probe analysis has shown that the *sapA-p* promoter of the spore-associated-protein (*sapA*) gene of *S. coelicolor* (51,52) is present only after growth in a colony has ceased, and S1 nuclease protection analyses of the *nshA-p1* promoter of *Streptomyces actuosus* (70) and of the *A1-p* promoter isolated from a *Streptomyces aureofaciens* plasmid (42) showed that these promoters are not active until late in the growth

**Table 5.** Analysis of 28 streptomycete promoter sequences similar to those recognized by RNA polymerases containing the $\sigma^{70}$ subunit in *Escherichia coli* and $\sigma^{A}$ in *Bacillus subtilis*.

| Description | Number of examples of nucleotide in position of hexamer given | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | −35 region | | | | | | | −10 region | | | | | |
| nt position | 1 | 2 | 3 | 4 | 5 | 6 | -spacer-[a] | 1 | 2 | 3 | 4 | 5 | 6 |
| T: | <u>25</u> | <u>26</u> | 0 | 2 | 2 | 5 | | <u>17</u> | 3 | 4 | 3 | 3 | <u>29</u> |
| A: | 0 | 0 | 0 | <u>20</u> | 3 | 8 | | 0 | <u>25</u> | 3 | 12 | 11 | 0 |
| G: | 4 | 3 | <u>29</u> | 1 | 5 | 10 | | 3 | 1 | 12 | 8 | 10 | 0 |
| C: | 0 | 0 | 0 | 6 | <u>19</u> | 6 | | 9 | 0 | 10 | 6 | 5 | 0 |
| Consensus: | T | T | G | A | C | (Pu) | | T | A | g | (Pu) | (Pu) | T |
| Percent: | 86% | 90% | 100% | 69% | 66% | 62% | | 59% | 86% | 41% | 69% | 72% | 100% |

[a]Three, 14, and 12 promoters had spacer regions of 16, 17, and 18 nucleotides, respectively.

curves of the respective organisms. These three temporally-regulated promoters do not share significant sequence similarities from −40 to +1 (Table 2).

### Mutational studies with streptomycete promoters

The promoters of four streptomycete genes (*ermE*, *aph*, *tsr*, *gal*), and one *E. coli* promoter expressed in *S. lividans* (*ampC-p*) have been analyzed by deletion or site-specific mutation analysis. The −35 region (TTGTCA) of the wild-type *E. coli ampC* promoter is separated from the −10 region (TACAAT) by 16 nucleotides (128). As mentioned previously, insertion of an additional nucleotide in the spacer region resulted in a 16-fold and 30-fold increase in promoter activity in *E. coli* and *S. lividans*, respectively (128). Site-specific mutations in the −10 (from TACAAT to TA<u>T</u>AAT) and −35 (from TTGTCA to TTG<u>A</u>CA) regions, making *ampC-p* identical to the consensus *E. coli* promoter sequence (17,18), resulted in higher transcription activities in both *E. coli* and *S. lividans*. It is perhaps notable that the effects were 2 to 3-fold greater in *E. coli* than in *S. lividans* (128). These increases in promoter activity, however, were significantly smaller than the increase caused by changing the spacer region from 16 to 17 nucleotides (128).

The streptomycete promoters *aph-p1* (43,46) and *ermE-p1* (35,43) are similar in that they both promote transcription at the translation initiation codon and they have similar $E\sigma^{70}$-like −10 regions (CATGAT and TAGGAT, respectively). Site-specific mutagenesis of the 3′ T (which is considered to be the most strongly conserved nucleotide of the −10 region; 17) in −10 region (to make CATGA<u>C</u> and TAGGA<u>C</u>, respectively) resulted in reduction of the transcript level below levels detectable by S1 nuclease protection analysis (35,43,45,46). Mutation of the −35 region of *ermE-p1* by site-directed mutagenesis of the TGGACA to <u>AAA</u>ACA resulted in no effect on the level of transcript obtained, and deletion of the TGG nucleotides (resulting in the formation of a −35 region of <u>GGC</u>ACA) resulted in slightly increased transcript levels (45). Thus, deletion of the TGG moved a potential, and presumably normally non-functional, conventional −35 region to a reasonable position upstream of the presumed −10 region. This indicates that the hexamer TGG-ACA, which actually resides around −28, is not important to the activity of the promoter.

Replacement of the sequences upstream from −22 of *tsr-p2*, resulting in two very different significant alterations in the nucleotide sequences of the −35 region, did not affect the location of the transcription start site of *tsr-p2* (47). On the other hand, recent experiments on *E. coli* promoters have shown that both the −10 and −35 consensus hexamers, along with the nucleotide sequences in the spacer region, are important for providing both specificity and efficiency of transcription (142,143). Data suggest that sequences upstream of −6 are not required for efficient promoter activity *in vitro* (142). Nevertheless, Bibb and Janssen's experiments (35,43−47), as a whole, suggest a minor role for the interaction of streptomycete RNA polymerases with the −35 regions of these particular promoters. Moreover, Baum et al. (38) isolated a DNA fragment from *Micromonospora echinospora* that contained multiple tandem promoters that were active in both *M. echinospora* and *S. lividans*. They showed that *mP1a* and *mP1b* were active despite deletion of all sequences up to −5 and −17, respectively (38). These data seem to corroborate the results obtained by Janssen and Bibb that certain actinomycete promoters may have reduced requirements for upstream sequences (e.g., −35 region). Promoter *mP1c*, on the other hand, had significantly reduced activity when sequences up to −34 were deleted (38). Various studies have shown that not all of the information required for promoter activity in *E. coli* is found in the −10 and −35 hexamers. For example, ideal −10 and −35 hexamers, separated by 17 nucleotides, have been found in *E. coli* sequences which do not exhibit promoter activity (144). In some cases, sequences upstream of the −35 hexamers (up to −160) (145) or downstream of −10 (as far as +20) (146) have been shown to have important effects on promoter strength. On the other hand, deletion of the −35 regions of certain *E. coli* promoters has little or no effect on transcription efficiency, similar to the results observed with the streptomycete promoters described above (147). Finally, the region protected by *E. coli* RNA polymerase typically ranges from −50 to ca. +10 (148) and sometimes can extend to −100 or to +20 (149), indicating that sequences important to promoter efficiency potentially could

reside anywhere within this large region (137). Nevertheless, conserved −10 and −35 hexameric sequences have been shown to have significant contributions to promoter activities in most cases (17,18,137,142,143).

Westpheling and her colleagues have recently analyzed the −35 region of the *gal-p1* promoter using site directed mutagenesis. This promoter has a hexameric sequence of GTGACA (similar to an *E. coli* $E\sigma^{70}$-like −35 sequence, albeit 23 nucleotides upstream of the putative −10 hexamer) centered around ca. −38 and a hexamer of GGGGGG centered around −32 (68). Alteration of nucleotides in the TTGTGA hexamer resulted in no significant change in promoter activity whereas mutation of nucleotides within the hexameric G residues resulted in promoter down mutations (J. Westpheling, personal communication). There are three significant aspects to these results: (i) this amply demonstrates the care that must be taken not to overemphasize similarities based on sequence comparisons alone; (ii) this corroborates the stringency of the spacer distance between the −10 and −35 hexamers observed in *E. coli* promoters (128,136,143); and (iii) these data suggest a new streptomycete promoter sequence which a specific RNA polymerase may recognize.

It is clear from the data presented herein that the work on streptomycete promoters and the RNA polymerases that recognize them is still in its infancy. Considerable effort needs to be made to show functionality to the apparent promoter sequences shown in this study. Given the complexity of streptomycete morphological and physiological differentiation and the apparent differences between regulation in streptomycetes versus other eubacteria, numerous RNA polymerase holoenzymes would also be expected to be involved in directing the transcription of genes of primary and secondary metabolism. At the 'International Symposium on Biology of Actinomycetes' recently held in Madison, WI, several advances in the research on streptomycete promoter sequences were described by a wide variety of research groups. For examples, several new promoter sequences were described, DNA footprint analyses by several groups have identified specific sequences to which putative repressors bind within the promoter regions, and in depth mutational analyses of several promoter sequences were described. As these data become published, it should become apparent that regulation of gene expression in streptomycetes is quite complex indeed.

## ACKNOWLEDGEMENTS

Work in my laboratory is supported on grants DCB-9019585 from the National Science Foundation and GM-43345 from the National Institutes of Health. I am indebted to Kathleen Kendrick, Robert Woodman, Li Yun, Mervyn Bibb, Kelly Brown, Tobias Kieser, Keith Chater, and Janet Westpheling for their critical comments on this manuscript. I also thank Gary Janssen, Mervyn Bibb, and Janet Westpheling for sharing unpublished data with me. Floppy disks containing these promoter tables (IBM compatible; Wordperfect 5.1), complete with references, will be provided upon request through regular mail or e-mail (Strohl.2@osu.edu).

## REFERENCES

1. Chater, K.F. (1984) In Losick, R. and Shapiro, L. (eds.), Microbial Development. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 89−115.
2. Mendez, C., Brana, A.F., Manzanal, M.B., and Hardisson, C. (1985) *Can. J. Microbiol.* **31**, 446−450.
3. Horinouchi, S. and Beppu, T. (1988) In Okami, Y., Beppu, T., and Ogawara, H. (eds.), Biology of Actinomycetes '88, Japan Scientific Societies Press, Tokyo, pp. 71−75.
4. Lawlor, E.J., Baylis, H.A., and Chater, K.F. (1987) *Genes Devel.* **1**, 1305−1310.
5. Chater, K.F., Bruton, C.J., Plaskitt, K.A., Buttner, M.J., Mèndez, C., and Helmann, J.D. 1989. *Cell* **59**, 133−143.
6. Adamidis, T., Riggle, P., and Champness, W. (1990) *J. Bacteriol.* **172**, 2962−2969.
7. Gibb, G.D., and Strohl, W.R. (1988) *Can. J. Microbiol.* **34**, 187−190.
8. Hopwood, D.A., Malpartida, F., and Chater, K.F. (1987) In Kleinkauf, H., von Döhren, H., Dornauer, H., and Nesemann, G. (eds.), *Regulation of Seconday Metabolite Formation*, Verlagsgesellschaft mbH, Weinheim, pp. 23−33.
9. Strohl, W.R., Bartel, P.L., Li, Y., Connors, N.C., and Woodman, R.H. (1991) *J. Industr. Microbiol.* **7**, 163−174.
10. Hopwood, D.A., Malpartida, F., Kieser, H.M., Ikeda, H., Duncan, J., Fujii, I., Rudd, B.A.M., Floss, H.G., and Omura, S. (1985) *Nature* **314**, 642−644.
11. Epp, J.K., M.L. Huber, J.R. Turner, Goodson, T., and Schoner, B.E. (1989) *Gene* **85**, 293−301.
12. Bartel, P.L., Zhu, C.-b., Lampel, J.S., Dosch, D.C., Connors, N.C., Strohl, W.R., Beale, J.M., Jr., and Floss, H.G. (1990) *J. Bacteriol.* **172**, 4816−4826.
13. Noack, D., Geuther, R., Tonew, M., Breitling, R., and Behnke, D. (1988) *Gene* **68**, 53−62.
14. Taguchi, S., Kumagai, I., Nakayama, J., Suzuki, A., and Miura, K.-i. (1989) *Bio/technol.* **7**, 1063−1066.
15. Bender, E., Koller, K.-P., and Engels, J.W. (1990) *Gene* **86**, 227−232.
16. Helmann, J.D., and Chamberlin, M.J. (1988) *Annu. Rev. Biochem.* **57**:839−872.
17. Hawley, D.K., and McClure, W.R. (1983) *Nucleic Acids Res.* **11**, 2237−2255.
18. Harley, C.B., and Reynolds, R.P. (1987) *Nucleic Acids Res.* **15**, 2343−2361.
19. Chater, K.F., and Hopwood, D.A. (1989) In Hopwood, D.A. and Chater, K.F. (eds.), Genetics of Bacterial Diversity. Academic Press, London, pp. 24−52 and 129−151.
20. Cowing, D.W., Bardwell, J.C.A., Craig, E.A., Woolford, C., Hendrix, R.W., and Gross, C.A. (1985) *Proc. Natl. Acad. Sci. (USA)* **82**, 2679−2683.
21. Helmann, J.D., and Chamberlin, M.J. (1987) *Proc. Natl. Acad. Sci. (USA)* **84**, 6422−6424.
22. Kennedy, C. (1989) In Hopwood, D.A., and Chater, K.F. (eds.), Genetics of Bacterial Diversity. Academic Press, London, pp. 107−127.
23. Collado-Vides, J., Magasanik, B., and Gralla, J.D. (1991). *Microbiol. Rev.* **55**:371−394.
24. Gralla, J.D. (1991) *Cell* **66**, 415−418.
25. Errington, J. (1991) *Proc. R. Soc. Lond.* **B 244**, 117−121.
26. Losick, R., and Pero, J. (1981) *Cell* **25**, 582−584.
27. Buttner, M.J. (1989) *Molec. Microbiol.* **3**, 1653−1659.
28. Doi, R.H. (1984) *Biotechnol. Genet. Engn. Rev.* **2**, 121−155.
29. Janssen, G.R., Bibb, M.J., Smith, C.P., Ward, J.M., Kieser, T., and Bibb, M.J. (1985) In Bonventre, P.F., Morello, J.A., Schlesinger, S., Silver, S.D., and Wu, H.C. (eds.), Microbiology-1985, American Society for Microbiology, Washington, D.C., pp. 392−396.
30. Hopwood, D.A., Bibb, M.J., Chater, K.F., Janssen, G.R., Malpartida, F., and Smith, P. (1986) In Booth, I.R. and Higgins, C.F. (eds.), Regulation of Gene Expression, 25 Years On. Symp. Soc. Gen. Microbiol., University of Cambridge Press, Cambridge, pp. 251−276.
31. Hutchinson, C.R. (1987) *Appl. Biochem. Biotechnol.* **16**, 169−189.
32. Hütter, R., and Eckhardt, T. (1988) In Goodfellow, M., Williams, S.T., and Mordarski, M. (eds.), Actinomycetes in Biotechnology. Academic Press, pp. 89−184.
33. Seno, E.T., and Baltz, R.H. (1989) In Shapiro, S. (ed.), Regulation of Secondary Metabolism in Actinomycetes, CRC Press, Inc., Boca Raton, pp. 1−48.
34. Westpheling, J., Ranes, M., and Losick, R. (1985) *Nature* **313**, 22−27.
35. Bibb, M.J., Janssen, G.R., and Ward, J.M. (1985) *Gene* **41**, E357−E368.
36. Buttner, M.J., and Brown, N.L. (1987) *Gene* **51**, 179−186.
37. Buttner, M.J., Fearnley, I.M., and Bibb, M.J. (1987) *Mol. Gen. Genet.* **209**, 101−109.
38. Baum, E.Z., Buttner, M.J., Lin, L.-S., and Rothstein, D.M. (1989) *J. Bacteriol.* **171**, 6503−6510.
39. Baylis, H.A., and Bibb, M.J. (1988) *Molec. Microbiol.* **2**, 569−579.
40. Sinclair, R., and Bibb, M.J. (1989) *Gene* **85**, 275−282.

41. Westpheling, J., and Brawner, M. (1989) *J. Bacteriol.* **171**, 1355–1361.
42. Farkasovsky, M., Kormanec, J., and Kollarova, M. (1991) *Biochim. Biophys. Acta* **1088**, 119–126.
43. Bibb, M.J., and Janssen, G.R. (1986) In Alacevic, M., Hranueli, D., and Toman, Z. (eds.), Proceedings of the Fifth International Symposium on the Genetics of Industrial Microorganisms. Ognjen Prica Printing Works, Karlovac, Yugoslavia, pp. 309–318.
44. Janssen, G.R., and Bibb, M.J. (1988) In Okami, Y., Beppu, T., and Ogawara, H. (eds.), *Biology of Actinomycetes '88*. Japan Scientific Societies Press, Tokyo, pp. 374–379.
45. Janssen, G.R., and Bibb, M.J. (1988) *Dev. Industr. Microbiol.* **29**, 89–96.
46. Janssen, G.R., Ward, J.M. and Bibb, M.J. (1989) *Genes & Devel.* **3**, 415–429.
47. Janssen, G.R., and Bibb, M.J. (1990) *Mol. Gen. Genet.* **221**, 339–346.
48. Buttner, M.J., and Brown, N.L. (1985) *J. Mol. Biol.* **185**, 177–188.
49. Jaurin, B., and Cohen, S.N. (1985) *Gene* **39**, 191–201.
50. Forsman, M., and Jaurin, B. (1987) *Mol. Gen. Genet.* **210**, 23–32.
51. Schauer, A., Ranes, M., Santamaria, R., Guijarro, J., Lawlor, E., Mendez, C., Chater, K., and Losick, R. (1988) *Science* **240**, 768–772.
52. Guijarro, J., Santamaria, R., Schauer, A., and Losick, R. (1988) *J. Bacteriol.* **170**, 1895–1901.
53. Virolle, M.-J., Long, C.M., Chang, S., and Bibb, M.J. (1988) *Gene* **74**, 321–324.
54. Ingram, C., Brawner, M., Youngman, P., and Westpheling, J. (1989) *J. Bacteriol.* **171**, 6617–6624.
55. Babcock, M.J., and Kendrick, K.E. (1990) *Gene* **95**, 57–63.
56. Nakano, M.M., Kikuchi, T., Mashiko, H., and Ogawara, H. (1989) *J. Antibiot.* **42**, 413–422.
57. Narva, K.E., and Feitelson, J.S. (1990) *J. Bacteriol.* **172**, 326–333.
58. Stein, D.S., Kendall, K.J., and Cohen, S.N. (1989) *J. Bacteriol.* **171**, 5768–5775.
59. Vögtli, M., and Hütter, R. (1987) *Molec. Gen. Genet.* **208**, 195–203.
60. Wong, H.C., Ting, Y., Lin, H.-C., Reichert, F., Myambo, K., Watt, K.W.K., Toy, P.L., and Drummond, R.J. 1991. *J. Bacteriol.* **173**, 6849–6858.
61. Bedford, D.J., Lewis, C.G., and Buttner, M.J. 1991. *Gene* **104**, 39–45.
62. Doyle, D., McDowall, K.J., Butler, M.J., and Hunter, I.S. 1991. *Mol. Microbiol.* **5**, 2923–2933.
63. Caballero, J.L., Malpartida, F., and Hopwood, D.A. 1991. *Mol. Gen. Genet.* **228**, 372–380.
64. Parro, V., Hopwood, D.A., Malpartida, F., and Mellado, P. 1991. *Nucleic Acids Res.* **19**, 2623–2627.
65. Hallam, S.E., Malpartida, F., and Hopwood, D.A. (1988) *Gene* **74**, 305–320.
66. Beck, C.F., and Warren, R.J. 1988. *Microbiol. Rev.* **52**, 318–326.
67. Neal, R.J., and Chater, K.F. (1991) *Gene* **100**, 75–83.
68. Fornwald, J.A., Schmidt, F.J., Adams, C.W., Rosenberg, M., and Brawner, M.E. (1987) *Proc. Natl. Acad. Sci. (USA)* **84**, 2130–2134.
69. Li, Y., Dosch, D.C., Strohl, W.R., and Floss, H.G. (1990) *Gene* **91**, 9–17.
70. Li, Y., Dosch, D.C., Woodman, R.H., Floss, H.G., and Strohl, W.R. 1991. *J. Industr. Microbiol.* **8**, 1–12.
71. Horinouchi, S., Suzuki, H., and Beppu, T. (1986) *J. Bacteriol.* **168**, 257–269.
72. Horinouchi, S., Kito, M., Nishiyama, M., Furuya, K., Hong, S.-K., Miyake, K., and Beppu, T. (1990) *Gene* **95**, 49–56.
73. Deng, Z., Kieser, T., and Hopwood, D.A. (1986) *Gene* **43**, 295–300.
74. Raibaud, A., Zalacain, M., Holt, T.G., Tizard, R., and Thompson, C.J. (1991) *J. Bacteriol.* **173**, 4454–4463.
75. Distler, J., Ebert, A., Mansouri, K., Pissowotzki, K., Stockmann, and Piepersberg, W. 1987. *Nucleic Acids Res.* **15**, 8041–8056.
76. Eckhardt, T., Strickler, J., Gorniak, L., Burnett, W.V., and Fare, L.R. (1987) *J. Bacteriol.* **169**, 4249–4256.
77. Smith, C.P., and Chater, K.F. (1988) *J. Mol. Biol.* **204**, 569–580.
78. Long, C.M., Virolle, M.-J., Chang, S.-Y., Chang, S., and Bibb, M.J. (1987) *J. Bacteriol.* **169**, 5745–5754.
79. Vigal, T., Gil, J.A., Daza, A., Garcia-Gonzalez, G., and Martin, J.F. (1991) *Mol. Gen. Genet.* **225**, 278–288.
80. Burnett, W.V., Henner, J., and Eckhardt, T. (1987) *Nucleic Acids Res.* **15**, 3926.
81. Taguchi, S., Nishiyama, K.-i., Kumagai, I., and Miura, K.-i. (1989) *Gene* **84**, 279–286.
82. Horii, M., Ishizaki, T., Paik, S.-Y., Manome, T., and Murooka, Y. (1990) *J. Bacteriol.* **172**, 3644–3653.
83. Kamimiya, S., and Weisblum, B. (1988) *J. Bacteriol.* **170**, 1800–1811.
84. Horinouchi, S., Nishiyama, N., Nakamura, A., and Beppu, T. (1987) *Mol. Gen. Genet.* **210**, 468–475.
85. Kovacevic, S., Tobin, M.B., and Miller, J.R. (1990) *J. Bacteriol.* **172**, 3952–3958.
86. Lopez-Cabrera, M., Perez-Gonzalez, J.A., Heinzel, P., Piepersberg, W., and Jiminez, A. (1989) *J. Bacteriol.* **171**, 321–328.
87. Chang, P.C., Kuo, T.-C., Tsugita, A., and Lee, Y.-H.W. (1990) *Gene* **88**, 87–95.
88. Holmes, D.J., and Cundliffe, E. 1991. *Mol. Gen. Genet.* **229**, 229–237.
89. Leu, W.-M., Wu, S.-Y., Lin, J.-J., Lo, S.J., and Lee, Y.-H.W. (1989) *Gene* **84**, 267–277.
90. Vujaklija, D., Ueda, K., Hong, S.-K., Beppu, T., and Horinouchi, S. 1991. *Mol. Gen. Genet.* **229**, 119–128.
91. Guilfoile, P.G., and Hutchinson, C.R. (1991) *Proc. Natl. Acad. Sci. (USA)* **88**, 8553–8557.
92. Smith, M.C.M., and Owen, C.E. 1991. *Mol. Microbiol.* **5**, 2833–2844.
93. Wray, L.V., Jr., and Fisher, S.H. (1988) *Gene* **71**, 247–256.
94. Murray, I.A., Gil, J.A., Hopwood, D.A., and Shaw, W.V. (1989) *Gene* **85**, 283–291.
95. Salauze, D., Perez-Gonzalez, J.A., Piepersberg, W., and Davies, J. (1991) *Gene* **101**, 143–148
96. Horinouchi, S., Furuya, K., Nishiyama, M., Suzuki, H., and Beppu, T. (1987) *J. Bacteriol.* **169**, 1929–1937.
97. Buttner, M.J., Chater, K.F., and Bibb, M.J. (1990) *J. Bacteriol.* **172**, 3367–3378.
98. Murakami, T., Holt, T.G., and Thompson, C.J. (1989) *J. Bacteriol.* **171**, 1459–1466.
99. Raymer, G., Willard, J.M.A., and Schottel, J.L. (1990) *J. Bacteriol.* **172**, 7020–7026.
100. Rosteck, P.R. Jr., Reynolds, P.A., and Hershberger, C.L. (1991) *Gene* **102**, 27–32.
101. Hoshiko, S., Makabe, O., Nojiri, C., Katsumata, K., Satoh, E., and Nagaoka, K. (1987) *J. Bacteriol.* **169**, 1029–1036.
102. Holmes, D.J., Drocourt, D., Tiraby, G., and Cundliffe, E. (1991) *Gene* **102**, 19–26.
103. Daza, A., Martin, J.F., and Gil, J.A. 1991. *Gene* **108**, 63–71.
104. Nakai, R., Horinouchi, S., and Beppu, T. 1988. *Gene* **65**, 229–238.
105. Rebollo, A., Gil, J.A., Liras, P., Asturias, J.A., and Martin, J.A. (1989) *Gene* **79**, 47–58.
106. Hoshiko, S., Nojiri, C., Matsunaga, K., Katsumata, K., Satoh, E., and Nagaoka, K. (1988) *Gene* **68**, 285–296.
107. Horinouchi, S., Suzuki, H., Nishiyama, M., and Beppu, T. (1989) *J. Bacteriol.* **171**, 1206–1210.
108. Horinouchi, S., Ishizuka, H., and Beppu, T. (1991) *Appl. Environ. Microbiol.* **57**, 1386–1393.
109. Zalacain, M., Gonzalez, A., Guerrero, M.C., Mattaliano, R.J., Malpartida, F., and Jimenez, A. (1986) *Nucleic Acids Res.* **14**, 1565–1581.
110. Pulido, D., Zalacain, M., and Jimenez, A. (1980) *Biochem. Biophys. Res. Comm.* **150**, 270–274.
111. Lacalle, R.A., Pulido, D., Vara, J., Zalacain, M., and Jimenez, A. (1989) *Gene* **79**, 375–380.
112. Urabe, H., Lenzini, M.V., Mukaide, M., Dusart, J., Nakano, M.M., Ghuysen, J.-M., and Ogawara, H. (1990) *J. Bacteriol.* **172**, 6427–6434.
113. Piron-Fraipont, C., Lenzini, M.V., Dusart, J., and Ghuysen, J.-M. (1990) *Mol. Gen. Genet.* **223**, 114–120.
114. Wasylyk, B., Wasylyk, C., Matthes, H., Wintzerith, M., and Chambon, P. (1983) *EMBO J.* **2**, 1605–1611.
115. Brosius, J., Cate, R.L., and Perlmutter, A.P. (1982) *J. Biol. Chem.* **257**, 9205–9210.
116. Goodfellow, M., and Cross, T. (1984) In Goodfellow, M., Mordarski, M., and Williams, S.T. (eds.), The Biology of the Actinomycetes, Academic Press, London, pp. 7–164.
117. Rosenberg, M., and Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353.
118. Piepersberg, W., Distler, J., Ebert, A., Heinzel, P., Mansouri, K., Kayer, G., and Pissowotzki, K. (1988) In Okami, Y., Beppu, T., and Ogawara, H. (eds.), Biology of Actinomycetes '88. Japan Scientific Societies Press, Tokyo, pp. 86–91.
119. Shine, J., and Dalgarno, L. (1975) *Nature (London)* **254**, 34–38.
120. Bibb, M.J., and Cohen, S.N. (1982) *Mol. Gen. Genet.* **187**, 265–277.
121. Bibb, M.J., Bibb, M.J., Ward, J.M., and Cohen, S.N. (1985) *Mol. Gen. Genet.* **199**, 26–36.
122. Moran, C.P., Jr., Lang, N., LeGrice, S.F.J., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J., and Losick, R. (1982) *Mol. Gen. Genet.* **186**, 339–346.
123. Tinoco, I., Jr., Borer, P.N., Dengler, B., Levine, B., Uhlenbeck, O.C., Crothers, D.M., and Gralla, J. (1973) *Nature (New Biol.)* **246**, 40–41.

124. Turner, D.H., Sugimoto, N., Jaeger, J.A., Longfellow, C.E., Freier, S.M., and Kierzek, R. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**, 123–133.
125. Shepard, H.M., Yelverton, E., and Goeddel, D.V. (1982) *DNA* **1**, 125–131.
126. McLaughlin, J.R., Murray, C.L., and Rabinowitz, J.C. (1981). *J. Biol. Chem.* **256**, 11283–11291.
127. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S., and Stormo, G. (1981) *Annu. Rev. Microbiol.* **35**, 365–403.
128. Jaurin, B., and Cohen, S.N. (1984). *Gene* **28**, 83–91.
129. Thompson, C.J., and Gray, G.S. (1983) *Proc. Natl. Acad. Sci. (USA)* **80**, 5190–5194.
130. Ptashne, M., Backman, K. Humayaun, M.Z., Jeffrey, A., Maurer, R. and Sauer, R.T. (1976) *Science* **194**, 156–161.
131. Klock, G., and Hillen, W. (1986) *J. Mol. Biol.* **189**, 633–641.
132. Dunn, R., McCoy, J., Simsek, M., Kajumdar, A., Chang, S.H., RajBhandry, U.L. and Khorana, H.G. (1981) *Proc. Natl. Acad. Sci. (USA)* **78**, 6744–6748.
133. Blanck, A., and Oesterhelt, D. (1987) *EMBO J.* **6**, 265–273.
134. Betlach, J., Friedman, J., Boyer, H.W., and Pfeifer, F. (1984) *Nucleic Acids Res.* **12**, 7949–7959.
135. Lopez, P., Martinez, S., Diaz, A., Espinosa, M., and Lacks, S.A. (1989) *J. Biol. Chem.* **264**, 4255–4263.
136. Stefano, J.E., and Gralla, J.D. (1982) *Proc. Natl. Acad. Sci. (USA)* **79**, 1069–1072.
137. Jacquet, M.-A., Ehrlich, R., and Reiss, C. (1989) *Nucleic Acids Res.* **17**, 2933–2945.
138. Siebenlist, U., Simpson, R.B., and Gilbert, W. (1980) *Cell* **20**, 269–281.
139. Gardella, T., Moyle, H., and Susskind, M.M. (1989) *J. Molec. Biol.* **206**, 579–590.
140. Siegele, D.A., Hu, J.C., Walter, W.A., and Gross, C.A. (1989) *J. Molec. Biol.* **206**, 591–603.
141. Zuber, P., Healy, J., Carter, H.L. III, Cutting, S., Moran, C.P. Jr., and Losick, R. (1989) *J. Molec. Biol.* **206**, 605–614.
142. Lorimer, D.D., Cao, J., and Revzin, A. (1990) *J. Mol. Biol.* **216**, 275–287.
143. Jacquet, M.-A., and Reiss, C. (1990) *Nucleic Acids Res.* **18**, 1137–1143.
144. Stormo, G.D., Schneider, T.D., Gold, L., and Ehrenfeucht, A. (1982) *Nucleic Acids Res.* **10**, 2997–3011.
145. Bossi, L., and Smith, Y.M. (1984) *Cell* **39**, 643–652.
146. Deutschle, U., Kammerer, W., Gentz, R., and Bujard, H. (1986) *EMBO J.* **4**, 2213–2222.
147. Okamoto, T., Sugimoto, K., Sugisaki, H., and Takanami, M. (1977) *Nucleic Acids Res.* **4**, 2213–2222.
148. Duval-Valentin, G., and Ehrlich, R. (1987) *Nucleic Acids Res.* **15**, 575–594.
149. Duval-Valentin, G., and Ehrlich, R. (1988) *Nucleic Acids Res.* **16**, 2031–2044.