Sigma elements are position-specific for many different yeast tRNA genes

Suzanne B.Sandmeyer, Virginia W.Bilanchone, Doni J.Clark, Paul Morcos, Georges F.Carle¹
and Garrett M.Brodeur^{1,2}

Department of Microbiology and Molecular Genetics, California College of Medicine, University of California, Irvine, CA 92717, ¹Department of Genetics and ²Department of Pediatrics, Washington University School of Medicine, St Louis, MO 63110, USA

Received August 4, 1987; Revised and Accepted January 4, 1988

ABSTRACT

We determined the DNA sequence of seventeen sigma elements and flanking regions in order to investigate the extent of the association between the yeast repetitive element, sigma, and tRNA genes. Fifteen of seventeen sigma elements analyzed begin at position -19 to -16 with respect to the 5' end of a tRNA-coding sequence. This region is close to the initiation point of tRNA gene transcription and contains a sequence which is modestly conserved for a number of tRNA genes. Two pairs of identical sigma elements occur as the long terminal repeats of a sequence which, together with flanking sigma elements, has the structural properties of a retrotransposon; this element has been named Ty3 (manuscript submitted). Hybridization analysis of yeast chromosomal DNA separated by orthogonal field alternation gel electrophoresis (OFAGE) showed that Ty3 and isolated sigma elements are distributed over many chromosomes in the yeast genome.

INTRODUCTION

The yeast genome contains three repeated elements—delta (1-3), tau (4), and sigma (5, 6)—which are several hundred basepairs (bp) in length. Delta, sigma, and tau share approximately ten short (6—11 bp) regions of sequence homology and may be related (4). These elements are frequently found in genomic regions which also include tRNA genes (4, 7-10). The most extensively studied of the three repeated elements is delta. Delta sequences are 330-338 bp in length and occur both as isolated elements and as the terminal direct repeats of the retrotransposons Ty1 and Ty2 (1-3, 11-13). Independent delta elements tend to exhibit some sequence heterogeneity and several clusters of these elements have been described (1, 10). These characteristics have made it difficult to estimate the total number of delta elements per typical haploid laboratory strain, but it is probably over 100 (1). Reversion of certain classes of Ty1 and Ty2 mutations results from recombination between the delta direct repeats (14, 15). This class of revertants is characterized by a single residual delta element. Thus, delta mobilization is apparently contingent on transposition of Ty1 and Ty2 elements, and isolated delta elements in the genome are thought to result from recombination between repeated delta elements.

Sigma is a highly conserved 340 or 341 bp element (5, 6) which is repeated about

30 times (16) in the haploid *Saccharomyces cerevisiae* genome. Although most sigma elements in common laboratory strains occur as isolated elements, they have also been found as the long terminal repeats of a longer retrovirus-like element, Ty3 (manuscript submitted). By analogy with the relationship between delta and the Ty1 and Ty2 elements, this finding suggests that the isolated sigma elements which have been characterized may be derived from recombination events between the sigma termini of Ty3 elements. The most striking feature of the sigma elements characterized to date is their very specific association with the 5' ends of tRNA genes (4-7, 9, 16-19). If sigma elements move(d) in Ty3-mediated transposition events, then the association between these elements and tRNA genes would suggest a novel position specificity for Ty3 as well.

The looser association previously noted of tau, delta, Ty1 and Ty2, with tRNA genes, suggests that tRNA genes offer target features for transposable elements in addition to Ty3(sigma). We undertook an extensive analysis of the sigma insertions in one strain of yeast in order to better understand the nature and extent of this targeting mechanism. Specifically, we examined seventeen sigma element loci from one strain of yeast in order to: 1) investigate the extent of sigma sequence conservation; 2) compare the sequences into which sigma has inserted; and 3) identify the tRNA genes with which sigma is associated. We present here the results of this analysis which show that sigma elements constitute a strictly conserved group of sequences which are closely associated with tRNA genes. Sigma elements exhibit a high degree of position specificity; they occur close to where tRNA gene transcription initiates, but not within a conserved "target" sequence.

MATERIALS AND METHODS

Reagents

Restriction endonucleases, DNA polymerase I (large fragment), and calf alkaline phosphatase were purchased from Boehringer Mannheim Biochemicals. T4 DNA ligase was obtained from New England Biolabs. Polynucleotide kinase was obtained from P-L Biochemicals. Radionucleotides, including [γ - 32 P]ATP (5000 Ci per mmol), [α - 32 P]dATP (3000 Ci per mmol), and [α - 32 P]dGTP (3000 Ci per mmol), were obtained from Amersham Corporation. The hexanucleotide primer mixture for randomly primed synthesis of radioactive probes was obtained from P-L Biochemicals. The plasmid pIBI20 was purchased from International Biotechnologies, Inc.

Cells and Culture Conditions

Cloning of genomic fragments from yeast strain AB972 into a Charon 30/lambda 1059 vector and identification of clones containing sigma elements was described previously (16). DNA was prepared from recombinant lambda phage grown in *E. coli* strain C600 (F⁻ thi-1 thr-1 leuB6 lacY1 tonA2 supE44). M13mp8 phage and pIBI20 plasmids

were cultured in JM109, (recA1 endA1 gyrA96 thi hsdR17 supE44 relA1 Δ [lac-proAB]) [F' traD36 proAB lacI^QZAM15] or JM103 (thi strA sbcB15 hsdR4 supE44 Δ [lac-proAB] [F' traD36 proAB lacI^QZAM15]) (20). Bacterial culture conditions were as described by Miller (21).

Yeast strains from which DNA was extracted (22) for Southern blot analysis (23) are AB972 (16) and YP148. YP148 has the genotype MAT α ura3-52 lys2 ade his7 trp1- Δ (personal communication, P. Hieter, Johns Hopkins University, Baltimore, Maryland). These strains were cultured by standard procedures (24).

Cloning Procedures and Sequence Analysis

Lambda clones 3, 7, 21, 23, 25, 26, 55, 84, 89, 148, and 161 were digested with Sau3A and were subcloned into the BamHI site of M13mp8. Phage containing each half of these sigma elements were identified by hybridization and their sequence determined by the dideoxynucleotide-chain termination method (25). The HindIII-EcoRI fragments from lambda clones 112 and 125, which contain sigma elements, were subcloned into the polylinker of the plasmid vector pIBI20, and the sequences of portions of these fragments were determined by the chemical modification method of Maxam and Gilbert (26). The strategy for sequence analysis of the 112 clone, which contains Ty3, is presented elsewhere (manuscript submitted). The sequence of the sigma-containing fragment from lambda clone 125 was determined from the XhoI site in the sigma element in both directions and from the SmaI site in the tRNA gene in both directions. The sigma element in clone 43 was subcloned into M13mp8 on a Sau3A fragment in each orientation and analyzed by the method of Sanger. The sigma element in lambda clone 55 was analyzed by a combination of methods. The tRNA gene-proximal portion of the sigma was subcloned into M13mp8 on a Sau3A fragment and the sequence was determined by the Sanger method. The 4.2 kbp HindIII-EcoRI fragment containing the entire sigma element and tRNA gene was subcloned into pIBI20. The sequence of this cloned fragment was determined from the XhoI site in both directions using the chemical-modification method. A HaeIII site occurs in the tRNA gene-proximal end of this sigma element. The sequence was determined from this site in both directions by the chemical-modification method, and this sequence overlapped sequence determined from the XhoI site. The sequence of the tRNA gene-proximal portion of the clone 134 sigma element was determined from a Sau3A fragment by the dideoxynucleotide-chain-termination method. The 0.85 kbp HindIII restriction fragment containing the sigma element plus the tRNA gene was also subcloned into the plasmid vector pIBI20. The sequence of the sigma element and the associated tRNA gene was determined by the chemical-modification method starting at the Sau3A site in the sigma element. The XhoI and Sau3A sites in sigma which were used in sequence analyses are indicated in Figures 1 and 2 in order to further clarify the strategy. All DNA sequence was compiled, compared, and edited using the DNA se-

quence analysis programs of Drs. A. Goldin (California Institute of Technology, Pasadena, California) and G. Gutman (University of California, Irvine, California).

Southern Blot Analysis of Chromosomal DNA

High molecular weight DNA was prepared from yeast strains AB972 and YP148 and fractionated in a 1.5% agarose gel by a modified OFAGE technique (22, 27,28, unpublished results). In the apparatus used, nearly homogeneous electric fields intersect at an angle of 115°. Electrophoresis was performed in 0.5 X TBE (1X = 0.09 M borate, 0.09 M Tris base, and 0.002 M Na₂H₂ EDTA [ethylene diamine tetracetic acid]) at 5 V/cm, with an 85 sec switching interval, for 40 h, at a temperature of 13°C. The separated chromosomal DNA was transferred to nitrocellulose by the method of Southern.

Hybridization conditions were as described (16). Hybridization probes consisted of gel-purified restriction fragments labeled by the random-primer method (29). The sigma probe was the XhoI-TaqI fragment from pPM55, a 339 bp fragment from 16 bp inside one end of sigma through 12 bp outside the other terminus. The Ty3 probe was a 1600 bp AccI fragment which is located between the two repeated sigma elements in lambda clone 112.

RESULTS

Comparison of the Sequences of Different Sigma Elements

Seventeen sigma element insertions were chosen for sequence analysis from a set of lambda clones containing inserts from the yeast strain AB972 (16). Ten sigma elements, for which the complete DNA sequence was determined, were compared in order to determine the extent of sequence conservation among the sigma elements of a single yeast strain (Fig. 1). Pairwise comparisons of these elements revealed no more than a 9-bp difference, or 2.6% sequence divergence, between any two elements. Six of the sigma elements, including the elements at both ends of the two Ty3 sequences, are identical. Figure 1 displays the five different sequences of the ten elements. Eight of the ten elements are 340 bp in length, although comparisons of the DNA sequences showed that some of these include single-bp deletions and compensating single-bp insertions. The DNA sequences of the sigma elements are ordered by degree of apparent divergence from the most common sequence. The apparent relatedness among elements, however, does not correlate with other common denominators, such as orientation, insertion target sequence, distance from the tRNA gene, or similarity of isoacceptors encoded by the tRNA gene.

Identification of tRNA Genes Associated with Different Sigma Elements

The sigma elements are located at a relatively fixed distance from the 5' ends of many different tRNA genes (Fig. 2). The tRNA genes associated with sigma elements in strain AB972 represent a broad spectrum of isoacceptor species, including two each for

Figure 1. Comparison of Sigma Sequences from Yeast Strain AB972. Complete sequence is shown for ten sigma elements from yeast strain AB972. These sigma elements are present in members of a charon 30/lambda 1059 yeast library in phages numbered 3, 7, 23, 26, 84, 112, 125, and 148. Two elements are contained in clone 112 and in clone 148. Sigma elements from clones 23, 84, 112, and 148 have the sequence shown on the line labeled 112. The sequences of sigma elements from clones 26, 3, 125, and 7 are shown in order of increasing divergence from that sequence. Gaps are introduced where necessary, in order to maintain alignment of homologous sequences. Positions at which the same base occurs as in the 112 sequence are indicated with a dash. The *Xho*I and *Sau*3A sites used in sequence analysis are labeled. The eight-bp terminal inverted repeats are indicated with arrows over the sequence. The strand shown is complementary to the transcript. The positions of transcription initiation in *MATa* cells treated with α -factor are indicated with asterisks. These are the same as described by Van Arsdell, Stetler and Thorner (19). Short interspersed regions of sequence which are conserved between sigma and delta elements (4) are shown in boldface type.

genes encoding isoacceptors for tRNA^{His}, tRNA^{Ala}, tRNA^{Arg}, and tRNA^{Ile} and one each for tRNA^{Lys}, tRNA^{Glu}, tRNA^{Ser}, tRNA^{Cys}, tRNA^{Tyr}, tRNA^{Asp}, and tRNA^{Val}.

The assignment of tRNA gene identities is discussed in the legend of Figure 3. The DNA which encodes the anticodons of these tRNA species is indicated in Figure 2 (right

Figure 2. Association of Sigma Elements and tRNA Genes in Yeast Strain AB972. The positions and orientations of 15 sigma elements and associated tRNA genes are shown. Elements are grouped by sigma orientation and arranged according to increasing lambda clone number. Left column, the family designation of the overlapping lambda clones (16); second column, the number of the lambda clone which was subcloned for sequence analysis; central column, the configuration of the sigma elements and the associated tRNA genes; and right column, the anticodon region of the tRNA gene. Repetitive elements sigma, delta, and tau, are indicated by lightly stippled, densely stippled, and hatched boxes, respectively. The restriction sites for XhoI (X) and Sau3A (S) within the sigma sequence are indicated for the topmost sigma element in each orientation. Sigma elements in the most common orientation are shown in the same orientation as the sequence presented in Figure 1. The tRNA genes are shown as open boxes. Cognate amino acids for the tRNAs are indicated, and in cases where the tRNA gene potentially encodes an identified tRNA isoacceptor species, that is also specified. See the legend of Figure 3 for a discussion of tRNA gene sequences. The number of basepairs between the terminal basepair of the sigma element and the initial basepair of the tRNA gene sequence is shown.

column). With the exception of the tRNA^{Ala}_{GCA} gene, these genes encode tRNAs which recognize frequently used codons in yeast (41, 42).

Position and Orientation of Sigma Element Insertions Relative to the Associated tRNA Gene

The previously observed bias in the insertion orientation of the sigma elements, relative to tRNA genes (16) is further supported by sequence analysis of these sigma

A

	-50	-45	-40	-35	-30	-25	-20	-15	-10	-5	1	5	10	15	20	25	30	35
GLU3 (λ 55)											TTTT	AGTTAATTTGTTTCCGATATAGTGTAA	GGCT	ATCACATACCGCTTTCA				
ASP ⁺ (λ 89)											ATATC	ACATCTAATATTCGGTGATAGTTTAAT	GGTC	AGAATGGGGCGTTTGTG				
ARG3A (λ 134)											CTTCC	AACAACAGTAGCTCGCGTGGCGTAAT	GGC	AACGGCTCTGACTTTCTA				
ILE (λ 161)											TCTTT	AACAAATAAATGGTCTCTTGGCCAGTTGGTT		AAGCCACCGTGGCTAAIA				
LYS2 (pFD17)*											CGCTC	CGAAGAACTCATCCITTTAGCTCAGTTTGGT		AGAGCTTCCGGCTTTTA				
LEU3 (pL120)*											ATTTC	AACAAATAAGTGTTCTTTGGCCGACG	GCTC	TAAAGCGCGCTGATTCAAG				
GLU3 (pY20)*											CATAC	GTGTATTGTAACTCGGATATAGTGTAA	GGCT	ATCACATACCGCTTTCA				
HIS (pYD2, λ 3)**	AATAGACAACACACACAAAAA	ACTGAAGGTCA	AAGTT	CATAAAGAAATTAGCCATCTT	AGTATAGT	GGTT	AGTACACATCGTTTGTG											
LYS1 (λ 23)	ATAGAGAATCTATTATCGATTTC	AACGGTGTAT	CAATC	ACGTTTAAAGCGCTT	TGGCCAACTCGT		AGCGCGTATGACTTTTA											
ALA (λ 26)	TTCCCAAAATGATTTACCACTCATT	AGAAATGA	GTTC	AACAGTCATCAGGGCACA	TGGCGCAGTTGGT		AGCGCGCTTCCCTTCCA											
SER2 (λ 84)	TTTTCTTAATCGTTTACATATTTT	ATCACTGAA	AATAA	CATACAATTTTGGCAACT	TGGCGCAGT	GGTT	AAGCGCAAGATTAGAA											
CYS (λ 112)	GTCAATTTCCCAATAATAGAGCAT	CAAAATGGAA	CTTTC	AGAAAATCAAGCTCGTAT	GGCGCAGT	GGT	AGCGCAGCAGATTGCAA											
TYR (λ 125)**	AGTTTGATTAATTAATAAGAAACAT	ATTTCAGC	CTTCG	CAAGAAATCGACTCTCGGT	AGCCCAAGTTGCTT		AAGCGCGCAAGCTGTAA											
SUQ5 (pPM15)*	TATTAATACCTTCGTTTCTTAACA	CTACC	AGATTATTTTGGCGACTAT	GGCCGAGT	GGTT		AAGCGCAAGCTGTAA											
SER2 (pY66)*	TAAGATTCTGATATTAGTCTTAC	ACTT	ATATAATAAGTGGCAACT	TGGCGCAGT	GGTT		AAGCGCAAGCTGTAA											
VAL (λ 43)											TAATC	AATAAGAATTTGGTTTCGGTGT	AGTCGGTT	ATGGCATCTGCTTAA				
ARG3A (λ 21)											CAAGT	GTAATTTCAATAGCTCGCGTGGCGTAAT	GGC	AACCGCTCTGACTTTCTA				
ALA (λ 7)	AACCTTATGGTATTCCTTATTTCT	TAATACCAC	TATTC	AACAAATATCAGGGCGTG	TGGCGTACTGCGGT		AGCGCGCTCCCTTACCA											
ILE (λ 148)	GTITGAGCACTGGGTATACGTTTT	TCCATATCATG	GAACA	TATGAAATATCTGGTCTCT	TGGCGCAGT	GGTT	AAGCGCACCGTGGCTAATA											
HIS (pYB12, λ 25)**	ATTTACACAAAAGTACAAAATAAG	GATACG	GTTC	CCATTGAAAAGTCGCCATCTT	AGTATAGT	GGTT	AGTACACATCGTTTGTG											

* Sequence determined in another study

** Same locus analyzed in another study and in this study

Figure 3. Sequences of Sigma Insertion Sites: the 5'-Flanking Regions of tRNA Genes.

A. Reconstructed sigma insertion sites. The sequences flanking sigma insertions, minus one copy of the 5-bp direct repeat, are shown here for twenty elements. Sequences are ordered by increasing lambda clone number and grouped by extent of known 5'-flanking sequence. The fifteen sequences flanking elements in the more common orientation are grouped on the top; the five sequences flanking elements in the reverse orientation are grouped on the bottom. The sequences in the 5' flanking region of tRNA^{Glu}, tRNA^{Asp}, tRNA^{Arg}, tRNA^{Ile}, tRNA^{His}, tRNA^{Lys}, tRNA^{Ala}, tRNA^{Ser}, tRNA^{Cys}, tRNA^{Tyr}, and tRNA^{Val} genes from strain AB972 were determined in this study. The SUQ5 tRNA^{Ser} gene sequence was reported earlier (6,31). Flanking sequences shown for tRNA^{His},

tRNA^{Lys} (5,30), tRNA^{Leu} (18), and tRNA^{Glu} (17) genes were previously reported by other workers. The plasmids or phage clones containing the sequences shown are indicated in parentheses. Positions in the 5'-flanking region of the tRNA-gene sequence are aligned and numbered, starting with -1. A space is introduced on either side of the 5-bp repeat. Conserved bases within the tRNA-coding region are shown in boldface type and nucleotides are numbered within the tRNA-coding region according to the yeast tRNA^{Phe} convention. Sequence is shown through the anticodon region.

tRNA Gene Assignments. The tRNA encoded by lambda clone 55 is homologous to tRNA^{Glu}₃ (32). Except for a mismatch in position -8, the gene appears to be allelic to the pY106 tRNA^{Glu} described by Eigel and Feldmann (7). The sequence determined from lambda clone 89 contains 48 bp which are homologous to the first 48 bp of tRNA^{Asp} determined by Gangloff et al. (33), but is not allelic to the tRNA^{Asp} genes analyzed by Schmidt et al. (34) and Gafner et al. (8). Lambda clones 134 and 21 contain sequence homology with the tRNA^{Arg}_{3A} sequence as determined by Kuntzel et al. (35). Sixteen bp of the 5'-flanking sequence are also identical to the tRNA^{Arg}_{3A} gene in plasmid pJB18U that Schmidt et al. (34) described, but the 3'-flanking sequences are not identical (134: CTTTTTTTGAATAAGCGTTGA; pJB18U: CTTTGTTTCCTCCGTG). Lambda clones 161 and 148 contain sequences which have perfect homology to the tRNA sequence determined by Pixa et al. (36) for tRNA^{Ile}. The two sequences determined in our study show A at positions encoding inosine in the tRNA. Lambda clones 3 and 25 contain the tRNA^{His} genes reported by del Rey et al. (5,30) in plasmids pYD2 and pYB12, respectively. Only the sequence of the tRNA gene-proximal portion of the sigma element in lambda clone 25 was determined, along with the tRNA gene in our study. Lambda clone 23 contains a tRNA gene with complete homology to the tRNA^{Lys} sequence determined by Smith et al. (37). Differences in 5'-flanking and coding regions indicate that this gene is distinct from the tRNA^{Lys} gene described by del Rey et al. (5), which is flanked by a sigma element. Lambda clone 7 has homology to the tRNA^{Ala} sequence reported by Penswick et al. (38) at all 73 positions. Lambda clone 26 is homologous to this sequence at 58 positions. The anticodons differ in the two genes: encoded by TGC in the case of lambda 26 and AGC in the case of lambda 7. Most of the differences between the two genes involve changes in unpaired regions or paired changes which preserve the stem configurations. One exceptional difference occurs at position 5 of the tRNA-coding sequence, which is G in lambda 7, and A (creating an AC pair in the amino acyl stem) in lambda 26. Analyses of the effects of mismatch in this region of the amino acyl stem suggest that this gene may not produce active, mature tRNA (39,40). The most common codon for ALA in six yeast genes surveyed by Bennetzen and Hall (41) and Ikemura (42) is GCU (68%), which is complementary to the anticodon of the tRNA encoded by lambda clone 7. The tRNA^{Ser}₂ gene in lambda clone 84 is completely homologous to the tRNA^{Ser}₂ sequence reported by Zachau et al. (43), but the flanking sequences did not show significant homology with tRNA^{Ser}₂ genes which have been previously analyzed (44,45). The tRNA^{Cys} gene associated with the Ty3 in lambda clone 112 is homologous to the tRNA^{Cys} sequence reported by Holness et al. (46), but is apparently not allelic to the tRNA^{Cys} gene associated with the tau element characterized by Genbauffe et al. (4). The tRNA^{Tyr} gene, SUP2, described here is allelic to the gene associated with a sigma sequence which was previously analyzed (6). The sigma element in lambda 43 is adjacent to sequence which is homologous, at least as far as the T loop, to the tRNA^{Val}₁ sequence reported from yeast by Bonnet et al. (47). This gene is not allelic to the tRNA^{Val}₁ gene described by Baker et al. (44).

B. Comparison of the 5'-flanking sequence of tRNA genes with and without sigma elements. DNA sequences from the 5'-flanking regions of twenty-six tRNA genes which do

not have sigma insertions were compared with DNA sequences from the 5'-flanking regions of twenty tRNA genes which do have sigma insertions. The positions are displayed from left to right, -16 to -1, relative to the mature tRNA-coding sequence. Solid bars are derived from sequences upstream of tRNA genes with sigma insertions; hatched bars are derived from sequences upstream of tRNA genes without sigma insertions. %G, A, T, or C in a particular position is shown in the panel indicated; the scale is from 0 to 80%. Sequences from the 5'-flanking region of genes for the following tRNAs were used in this analysis: tRNA^{Ser} (SUP-RL1;31); tRNA^{Ser} (B, G, J;45); tRNA^{Ser} (SUP19, 17, 16;48); tRNA^{Met} (PYMT-5;49) tRNA^{Met} (I and II;50) tRNA^{Glu}₃ (pY5;51); tRNA^{Glu}₃ (pY106;7); tRNA^{Gln} (pY66;9); tRNA^{Phe} (pYPT2 and pYPT5;52); tRNA^{Leu}; (SUP53;53); tRNA^{Leu}₃ (54); tRNA^{Arg}₂ (pY41;44) tRNA^{Arg}₃ (PJB-18U and PJB-19F;34); tRNA^{Asp} (H13;8); tRNA^{His} (pYG2;30) and tRNA^{Tyr} (SUPP4, 11, and 8;55).

insertions. Ten of the fifteen elements are in the same orientation, relative to the associated tRNA gene. Although the sigma direct repeats of the two Ty3 elements are in the same orientation relative to the internal domain, the two Ty3 elements in AB972 are in opposite orientations, relative to the associated tRNA gene. Orientation of sigma elements does not appear to correlate with a particular position of insertion. It is, nonetheless, a parameter which may have implications for the mechanism of insertion, and for the effect of the insertion on tRNA gene expression. Because the direction of transcription of the sigma element is distal to proximal, relative to the *Xho*I site (manuscript submitted; 19), transcription of elements, in the more common orientation, is away from the associated tRNA gene and, in the less common orientation, toward the tRNA gene. In the case of the Ty3-terminal sigma elements, transcription is from the left sigma element toward the internal domain (manuscript submitted).

Fifteen of the seventeen sigma elements analyzed in this study were closely associated with tRNA genes. Eight of the fifteen elements are at position -17, relative to the 5' end of the tRNA coding sequence; three occur at position -18; two occur at position -16; and two occur at position -19 with respect to the 5' end of mature tRNA-coding sequence. The remaining two elements are the tRNA gene-distal long terminal repeats of two Ty3 elements, and so are approximately 5.1 kilobase pairs (kbp) from the tRNA gene sequence. Figure 2 shows a summary of the positions and orientations of the fifteen sigma elements, relative to the associated tRNA genes. The DNA sequences in the regions flanking the sigma elements are displayed in Figure 3A.

A comparison of the sequences of the insertion sites of seventeen sigma elements determined in this study and the insertion sites of several sigma elements which were previously described was carried out in order to determine whether similarities exist near the site of insertion (Fig. 3). Thirteen of these twenty insertions begin at position -17, relative to the 5' end of the tRNA coding sequence. Insertion at the position of staggered cleavage is inferred from the presence of the 5-bp direct repeats which flank

Figure 4. Chromosomal Distribution of Sigma and Ty3 Elements. A and C, ethidium bromide-stained chromosomal DNA from yeast strains AB972 and YP148, respectively, separated by OFAGE. B and D, the chromosomal DNA shown in A and C, transferred and hybridized to a radiolabeled sigma-specific probe (side 1) or Ty3 internal domain-specific probe (side 2). Chromosome assignments are shown between panels. These assignments were based on position of migration and hybridization to chromosome-specific probes (57; unpublished observations).

sigma insertions. In nine out of the twenty sigma element insertion sites compared, the tRNA gene-proximal nick would have occurred within the sequence $5'TCA3'$ $3'AGT5'$ and at seven additional insertion sites, this nick would have occurred before, after, or within the dinucleotide bp $5'CA3'$ $3'GT5'$ or $5'TC3'$ $3'AG5'$ (Fig. 3A).

In order to determine whether this limited homology distinguishes the 5'-flanking region of tRNA genes which have sigma insertions from those which do not, these sequences were compared with the sequences of the 5'-flanking regions of 26 tRNA genes which are not associated with sigma elements. The result of this comparison is shown in Figure 3B. Eighteen of these 26 tRNA gene-flanking sequences had the trinucleotide bp 5'TCA3' 3'AGT5' between positions -10 and -15. The sequence TTTCAACAAATAAGT, which includes the TCA motif, was previously observed in the 5'-flanking regions of fifteen tRNA genes (56). Thus, this sequence apparently does not distinguish tRNA genes which have sigma insertions from those which do not.

Chromosomal Distribution of Sigma and Ty3 Elements

The chromosomal distribution of Ty3 elements and isolated sigma elements was examined in two haploid yeast strains by separation of yeast chromosomes using OFAGE analysis of the DNA on Southern blots with sigma-specific and Ty3 internal domain-specific DNA probes (Fig. 4). The two strains examined are AB972, the strain from which the analyzed sigma elements were cloned (Fig. 4, A and B) and YP148, a strain in which fragmentation of chromosome VII allows resolution of chromosomes VII and XV and a chromosome-length polymorphism allows separation of chromosomes V and VIII (Fig. 4, C and D). The chromosomes were identified by hybridization to chromosome-specific probes, as previously described by Carle and Olson (57, unpublished data). The sigma probe did not hybridize to chromosome VI or XIII in either strain, nor did it hybridize to chromosome fragment VIIA from strain YP148. Chromosomes VII + XV and V + VIII from strain AB972 are electrophoretically indistinguishable, and each pair showed intense hybridization to sigma sequences. Exposures of each autoradiograph were scanned in the linear range with a laser densitometer (LKB UltroScan XL), and the area under each peak was integrated. In each sample, chromosome I hybridized distinctly to the sigma probe, but was among the most weakly hybridizing of the chromosomes. If chromosome I is assumed to have one sigma element, and the number of sigma elements on the other chromosomes is estimated by relative band intensity, then 34 and 31 are lower limits for the number of elements in strains AB972 and YP148, respectively. This agrees well with the estimate of about 30 sigma elements for the AB972 genome, which was based on Southern blot analysis of EcoRI-HindIII digested genomic DNA (16).

In order to examine these strains for Ty3 insertion polymorphisms, a Ty3-specific DNA probe was hybridized to intact chromosomal DNA from strain AB972 and from strain YP148, separated by OFAGE. Under these conditions, the Ty3 internal domain does not cross hybridize with probes from Ty1 or Ty2 epsilon sequences. In the sample from AB972, the probe hybridized to DNA migrating in bands corresponding to chromosomes VII + XV, and IX. This agrees with Southern blot analyses of DNA digested with restriction enzymes which argue that only two copies of Ty3 exist in AB972 (manuscript

submitted). The same probe hybridized to chromosomes XII, VIIB, XI and IX from strain YP148, which suggests that a minimum of four Ty3 elements exist in the genome of this strain. Thus, Ty3 is present in different numbers and positions in the genomes of different yeast strains.

DISCUSSION

Positions of Sigma Insertions

The association observed between sigma and tRNA genes is extended by this sequence analysis of a set of sigma elements from one strain of yeast. This analysis showed that sigma elements are typically associated with genes for isoacceptors which decode frequently used codons. Comparisons of these sigma element sequences also showed that the element is highly conserved relative to another yeast repetitive element, delta. The most striking feature of the association between sigma and tRNA genes, however, is the positional specificity of the insertions. Out of the seventeen sigma elements analyzed, fifteen are separated by 19 bp or less from the 5' end of mature tRNA-coding sequences; the remaining two elements are the distal members of tRNA gene-associated sigma composite elements.

Although the significance of the association between highly conserved sigma elements and genes for isoacceptors which decode frequently used codons is not clear, our sequence analysis of these elements in one strain of yeast does suggest several possibilities. In yeast, frequently used codons have been shown to be decoded, in general, by major isoacceptor species (41, 42) and the available data supports the notion that major isoacceptors are encoded by the more redundant gene families (58). Thus, the location of sigma elements adjacent to these genes would be favored statistically. If insertion of Ty3(sigma) is detrimental to tRNA gene activity, insertions neighboring members of these larger gene families might also be more likely to be tolerated. Finally, if tRNAs themselves are in some way responsible for the position specificity, this insertion pattern, which mirrors the distribution of tRNA species in the cells, would be expected.

Conservation of the Sigma Sequence

The most simple explanation of the existence of isolated sigma elements is that, like solo delta elements, they arise from recombination between the long terminal repeats of a parent transposable element. In contrast to delta, however, the conservation of these sigma elements is quite striking. Conservation of repetitive sequences can be explained by recent proliferation, strong functional selection, and/or sufficient gene conversion to suppress evolutionary divergence. Although the first possibility would seem to be favored by the existence of relatively few Ty3 and sigma sequences compared to Ty1 and Ty2 and delta sequences, we recorded three instances of sigma elements interrupted by delta or tau termini. This observation, while not ruling out recent spread

of the sigma element as the basis for the observed sequence conservation, is not particularly supportive. Another explanation of sigma sequence conservation is that it confers some as yet unidentified selective advantage on the cell. Alternatively, if isolated sigma elements can transpose, then sigma sequence conservation could reflect the proliferation of a transpositionally active version of the element. This conservation could be the result of gene conversion. Local regions having high frequencies of gene conversion have been reported in the organisms where it has been possible to measure gene conversion accurately (59, 60). Of particular interest with respect to our analyses, are the previous observations of Amstutz et al. (61) that at least some tRNA^{Ser} genes in Saccharomyces pombe undergo gene conversion at a rate of at least 10^{-4} per meiosis. If gene conversion were generally elevated in the vicinity of tRNA genes, it might contribute to the maintenance of sequence uniformity among sigma elements.

Ty3 (Sigma) Insertion Specificity

One of the most intriguing puzzles posed by the close association between sigma elements and tRNA genes is the mechanism of target-site recognition by the element. Ty3 has a retrovirus-like organization of its coding sequences and encodes proteins with retrovirus homologues (Haywood, Chalker, and Sandmeyer, unpublished data). Ty3, like other retrotransposons and retroviruses (retroid elements) is flanked by short, direct repeats generated during integration. Because the length of the short, direct repeat flanking these insertions is characteristic of the retrovirus, rather than the host species (62), the endonuclease encoded by the retrovirus, rather than a host nuclease, is inferred to cleave the target sequence. Nevertheless, no data address this point directly. Positions of retroviral insertion have, in some cases, been correlated with DNase hypersensitive sites (63, 64), but not with any specific sequence. Similarly, inspection of reconstructed target sequences in our study did not reveal a sequence at the immediate target site which might be required for insertion. The sequence 5'TCA3' was observed in the tRNA-like strand upstream of almost half of the tRNA genes associated with sigma, but this sequence also occurs at a similar frequency upstream of tRNA genes not associated with the element. Therefore, its presence apparently does not increase the probability of a Ty3(sigma) insertion. The most simple interpretation of our data is that the position specificity of the sigma element is not solely based on the presence of a specific DNA sequence at the primary target (integration) site. Rather, we conclude that recognition must be based on: 1) some characteristic pattern or structure implicit in the tDNA sequence, 2) interactions between the integration apparatus and proteins complexed to the tRNA gene, or 3) some combination of these. We discuss below some features of tRNA genes which may bear on this unusual specificity.

The first proposed explanation of Ty3(sigma) insertion specificity—that the target is implicit in the DNA sequence—implies an integration apparatus which recognizes a

conserved sequence pattern (65) or a characteristic conformation of tDNA and cleaves upstream in the DNA. This idea is appealing because of its inherent simplicity, and also because polymerase III transcription of different tRNA genes must require similar gene-family recognition. It must be noted, however, that while the sigma elements are closely associated with the tRNA gene, the position of the insertion relative to beginning of tRNA-coding sequence or conserved sequences within the gene, is not absolutely constant. Thus, a DNA model would entail precise spacing of the insertion from some secondary or tertiary pattern not obvious in the primary sequence or must allow for some small variation in the distance of the insertion site from a conserved target in the tDNA.

The second proposed explanation of Ty3(sigma) insertion specificity involves tRNA gene-associated proteins in targeting. Two types of proteins could be related to tRNA gene targeting by sigma. First, as mentioned above, tRNA genes in yeast have been shown to undergo gene conversion at high frequencies, suggesting that they are predisposed to local recombination events. Proteins involved in these reactions could also make the DNA more vulnerable to nonhomologous recombination events, such as Ty3 insertion. Second, it is known that tRNA transcription must initiate very close to the site of sigma insertion. In fact, transcription initiation occurs in the sequence PyPyCAACAA, at the second A, for several tRNA genes (66; D. Engelke, University of Michigan, Ann Arbor, Michigan, personal communication), and Raymond et al. (56) observed a related sequence which affected the level of transcription of a tRNA^{Leu} gene and occurs in the 5'-flanking sequence of a number of tRNA genes. At least one transcription factor has been shown to interact with this region; tau, a polymerase III transcription factor required for tRNA gene transcription was shown to protect the flanking sequence of a tRNA^{Glu}₃ gene up to position -11 from weak DNase digestion (67). (In our study, tRNA gene-proximal cleavage would have occurred most frequently between positions -11 and -12.) Finally, electron microscopic observation of a tRNA gene suggests that the active gene may have a distinctive structure. Stillman et al. (68) found that tRNA genes bound to a transcription factor appeared bent, possibly due to contact of the factor with a bipartite internal promoter. Proteins which allow the DNA to be transcribed, then, are known to act on the DNA in this region and might mark it in some way as a target for transposition.

If the polymerase III transcription complex is central to Ty3(sigma) target recognition, then the 5S rRNA genes which are also transcribed by polymerase III might also be targets for insertion of this element. Yeast chromosome XII contains more than 100 tandem repeats of the rRNA genes, including those for 5S rRNA (69). According to the intensity of the band on the autoradiograph (Fig.4) corresponding to chromosome XII, however, this chromosome contains only two or three copies of sigma and none of Ty3. The distribution of these elements, then, does not reflect the distribution of all genes

transcribed by polymerase III, and this study provides no evidence that the repeated 5S genes are a hotspot for Ty3 insertion, as might be predicted if any polymerase III transcription complex constitutes a target. This observation is consistent with the target involving some tRNA gene-specific features. Nevertheless, this description of insertions cannot completely rule out a general association between Ty3(sigma) and polymerase III-transcribed genes. It is possible that Ty3(sigma) elements inserted near 5S rRNA genes are eliminated by unequal crossing-over among the tandemly repeated rRNA genes.

The suggestion of an association between transcription and recombination is not new. Polymerase I transcription activity is required for HOT1 activation of recombination in yeast (70). In addition, polymerase II activity is required for gene conversion of the MAT locus of S. cerevisiae (71) and recombination of a mouse immunoglobulin variable region has been shown to be enhanced by transcription (72). Although no other retroid element is known to have specificity of the type displayed by Ty3, genes transcribed by polymerase III are associated with other types of retroid elements. In yeast, Ty1 and Ty2 occur frequently within several hundred bp of the 5' ends of tRNA genes (7-10), and in animal cells, Alu sequences are both targets and donors in retro-transposition (73).

In conclusion, the extreme position specificity of the Ty3(sigma) insertion pattern is an intriguing natural curiosity. Further studies of Ty3(sigma) transposition will provide a potentially useful paradigm for understanding the mechanism of genomic integration of retroid elements.

Acknowledgments

This work was supported by National Institutes of Health Grant GM33281 to S.B.S.

REFERENCES

1. Cameron, J. R., Loh, E. Y., and Davis, R. W. (1979) *Cell* **16**, 739-751.
2. Farabaugh, P. J. and Fink, G. R. (1980) *Nature* **286**, 352-356.
3. Gafner, J. and Philippsen, P. (1980) *Nature* **286**, 414-418.
4. Genbauffe, F. S., Chisholm, G. E., and Cooper, T. G. (1984) *J. Biol. Chem.* **259**, 10518-10525.
5. del Rey, F. J., Donahue, T. F., and Fink, G. R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4138-4142.
6. Sandmeyer, S. B. and Olson, M. V. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7674-7678.
7. Eigel, A. and Feldmann, H. (1982) *EMBO J.* **1**, 1245-1250.
8. Gafner, J., DeRobertis, E. M., and Philippsen, P. (1983) *EMBO J.* **2**, 583-591.
9. Nelbock, P., Stucka, R., and Feldmann, H. (1985) *Biol. Chem. Hoppe Seyler* **366**, 1041-1051.
10. Warmington, J. R., Anwar, R., Newlon, C. S., Waring, R. B., Davis, R. W., Indge, K. J., and Oliver, S. G. (1986) *Nucleic Acids Res.* **14**, 3475-3485.

11. Kingsman, A. J., Gimlich, R. L., Clarke, L., Chenault, A. C., and Carbon, J. (1981) *J. Mol. Biol.* **145**, 619-632.
12. Roeder, G. S. and Fink, G. R. (1983) In Shapiro, J. A. (ed), Mobile Genetic Elements, Academic Press, New York, pp. 299-328.
13. Williamson, V. M. (1983) *Int. Rev. of Cytol.* **83**, 1-25.
14. Roeder, G. S. and Fink, G. R. (1980) *Cell* **21**, 239-249.
15. Chaleff, D. T. and Fink, G. R. (1980) *Cell* **21**, 227-237.
16. Brodeur, G. M., Sandmeyer, S. B., and Olson, M. V. (1983) **80**, 3292-3296.
17. Feldmann, H., Olah, J., and Friedenreich, H. (1981) *Nucleic Acids Res.* **9**, 2949-2959.
18. Standring, D. N., Venegas, A., and Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5963-5967.
19. Van Arsdel, S. W., Stetler, G. L., and Thorner, J. (1987) *J. Mol. Cell. Biol.* **7**, 749-759.
20. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**, 103-119.
21. Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
22. Carle, G. F. and Olson, M. V. (1984) *Nucleic Acids Res.* **12**, 5647-5664.
23. Southern, E. M. (1979) *Methods Enzymol.* **68**, 152-176.
24. Sherman, F., Fink, G. R., and Hicks, J. B. (1982) Methods in Yeast Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
25. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
26. Maxam, A. M. and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.
27. Schwartz, D. C., Saffran, W., Welsh, J., Haas, R., Goldenberg, M., and Cantor, C. R. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 189-195.
28. Carle, G. F., Frank, M., and Olson, M. (1986) *Science* **232**, 65-68.
29. Feinberg, A. P. and Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266-267.
30. del Rey, F., Donahue, T. F., and Fink, G. R. (1983) *J. Biol. Chem.* **258**, 8175-8182.
31. Olson, M. V., Page, G. S., Senenac, A., Piper, P. W., Worthington, M., Weiss, R. B., and Hall, B. D. (1981) *Nature* **291**, 464-468.
32. Kobayashi, T., Irie, T., Yoshida, M., Takeishi, K., and Ukita, T. (1974) *Biochim. Biophys. Acta* **366**, 168-181.
33. Gangloff, J., Keith, G., Ebel, J.-P., and Dirheimer, G. (1972) *Biochim. Biophys. Acta* **259**, 210-222.
34. Schmidt, O., Mao, J., Ogden, R., Bukmann, J., Sakano, H., Abelson, J., and Soll, D. (1980) *Nature* **287**, 750-752.
35. Kuntzel, B., Weissenbach, J., and Dirheimer, G. (1974) *Biochimie* **56**, 1069-1087.
36. Pixa, G., Dirheimer, G., and Keith, G. (1984) *Biochem. Biophys. Res. Commun.* **119**, 905-912.
37. Smith, C. J., Teh H.-S., Ley, A. N., and D'Obrean, P. (1973) *J. Biol. Chem.* **248**, 4475-4485.
38. Penswick, J. R., Martin, R., Dirheimer, G. (1975) *FEBS Lett.* **50**, 28-31.
39. Kurjan, J. and Hall, B. D. (1982) *Mol. Cell. Biol.* **2**, 1501-1513.
40. Willis, I., Hoffinger, H., Pearson, D., Chisholm, V., Leupold, U., and Soll, D. (1984) *EMBO J.* **3**, 1573-1580.
41. Bennetzen, J. L. and Hall, B. D. (1982) *J. Biol. Chem.* **257**, 3026-3031.
42. Ikemura, T. (1982) *J. Mol. Biol.* **158**, 573-597.
43. Zachau, H. G., Duetting, D., and Feldmann, H. (1966) *Hoppe-Selyer's Zeit. Physiol. Chem.* **347**, 212-235.

-
44. Baker, R. E., Eigel, A., Vogtel, D., and Feldmann, H. (1982) *EMBO J.* **1**, 291-295.
 45. Page, G. S. and Hall, B. D. (1981) *Nucleic Acids Res.* **9**, 921-934.
 46. Holness, N. J. and Atfield, G. (1976) *Biochem. J.* **153**, 447-454.
 47. Bonnet, J., Ebel, J. P., Dirheimer, G., Shershneva, L. P., Krutilina, A. I., Venkstern, T. V., and Bayev, A. A. (1974) *Biochimie* **56**, 1211-1213.
 48. Broach, J. R. (1981) *J. Mol. Biol.* (1981) **150**, 375-387.
 49. Venegas, A., Gonzalez, E., Bull, P., and Valenzuela, P. (1982) *Nucleic Acids Res.* **10**, 1093-1096.
 50. Cigan, A. M. and Donahue, T. F. (1986) *Gene* **41**, 343-348.
 51. Eigel, A., Olah, J., and Feldmann, H. (1981) *Nucleic Acids Res.* **9**, 2961-2970.
 52. Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R., and Rutter, W. J. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 190-194.
 53. Andreadis, A., Hsu, Y.-P., Kohlhaw, G. B., and Schimmel, P. (1982) *Cell* **31**, 319-325.
 54. Kang, H. S., Ogden, R. C., and Abelson, J. (1980) In *Twelfth Miami Winter Symposium*, pp. 317-333.
 55. Goodman, H. M., Olson, M. V., and Hall, B. D. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5453-5457.
 56. Raymond, K. C., Raymond, G. J., and Johnson, J. D. (1985) *EMBO J.* **4**, 2649-2656.
 57. Carle, G. F. and Olson, M. V. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3756-3760.
 58. Guthrie, C., and Abelson, J. (1982) In *Strathern, J. N., Jones, E. W., and Broach, J. R. (eds.) The Molecular Biology of the Yeast Saccharomyces, Metabolism and Gene Expression*, Cold Spring Harbor Laboratory, NY, pp. 487-560.
 59. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983) *Cell* **33**, 25-35.
 60. Whitehouse, H. L. K. (1983) *Nature* **306**, 645-646.
 61. Amstutz, H., Munz, P., Heyer, W. D., Leupold, U., and Kohli, J. (1985) *Cell* **40**, 879-886.
 62. Varmus, H. E. (1983) In *Shapiro, J. A. (ed.) Mobile Genetic Elements*, Academic Press, New York, NY, pp. 411-503.
 63. Robinson, H. L., and Gagnon, G. C. (1986) *J. Virol.* **57**, 28-36.
 64. Vijaya, S., Steffen, D. L., and Robinson, H. L. (1986) *J. Virol.* **60**, 683-692.
 65. Galli, G., Hofstetter, H., and Birnstiel, M. L. (1981) *Nature* **294**, 626-631.
 66. Engelke, D. R., Gegenheimer, P., and Abelson, J. (1985) *J. Biol. Chem.* **260**, 1271-1279.
 67. Camier, S., Gabrielsen, O., Baker, R., and Sentenac, A. (1985) *EMBO J.* **4**, 491-500.
 68. Stillman, D. J., Better, M., and Geiduschek, E. P. (1985) *J. Mol. Biol.* **185**, 451-455.
 69. Petes, T. D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 410-414.
 70. Voelkel-Meiman, K., Keil, R. L. and Roeder, G. S. (1987) *Cell* **48**, 1071-1079.
 71. Klar, A. J. S., Strathern, J. N., and Hicks, J. B. (1981) *Cell* **25**, 517-524.
 72. Blackwell, T. K., Moore, M. W., Yancopoulos, G. D., Suh, H., Lutzker, S., Selsing, E., and Alt, F. W. (1986) *Nature* **324**, 585-589.
 73. Weiner, A. M., Deininger, P. L., and Efstratiadis, A. (1986) *Ann. Rev. Biochem.* **55**, 637-661.
-