Nucleic Acids Research, 1998, Vol. 26, No. 3 689–696

tRNA genes and retroelements in the yeast genome

Jean Hani and Horst Feldmann[1,*]

Munich Information Centre for Protein Sequences, Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany and [1]Adolf-Butenandt-Institut für Physiologische Chemie der Universität München, Schillerstrasse 44, D-80336 München, Germany

Received October 1, 1997; Revised and Accepted November 20, 1997

ABSTRACT

A survey of tRNA genes and retroelements (Ty) in the genome of the yeast *Saccharomyces cerevisiae* is presented. Aspects of genomic organization and evolution of these genetic entities and their interplay are discussed. Attention is also given to the relationship between tRNA gene multiplicity and codon selection in yeast and the role of Ty elements.

INTRODUCTION

The yeast *Saccharomyces cerevisiae* is the first eukaryotic organism for which the entire genomic sequence has been determined (1,2). This information not only represents an essential 'navigational aid' for studies of other eukaryotic genomes, it will serve as a standard for genome comparisons throughout (3). A wealth of data is available for some 6000 open reading frames (ORFs) (4). Detailed information has also been obtained on other genetic entities, such as the genes for the various RNA species and yeast retrotransposons, the Ty elements.

The unique position of *S.cerevisiae* as a model eukaryote owes much to its intrinsic advantages as an experimental system. It nearly goes without saying that yeast has served, for example, as a model organism and convenient source of material in the analysis of eukaryotic tRNAs and tRNA genes (for a review see 5) and as a useful system for understanding interactions between retroelements and their host (6).

At a first sight the (repetitious) genetic entities specifying tRNAs and Ty elements seem to have little in common. However, recent work has clearly established that tRNA genes and tRNAs influence several key steps in transposition (7,8). The close association of the majority of Ty elements with tRNA genes (9) and interactions between Ty elements and tRNAs also reflect the intricate ways in which these transposable elements and their host must have co-evolved (7–9).

We have compiled a catalogue of yeast tRNA genes and Ty elements and data pertinent to their abundance and genomic distribution. This information has been used to revisit interesting aspects, such as the relationship between tRNA gene multiplicity and codon selection in yeast and the significance of Ty elements. Finally, we will address questions as to the evolution of these genetic entities.

YEAST tRNAs AND tRNA GENES

Number and genomic distribution of yeast tRNA genes

During the course of the yeast sequencing project we relied on computer-aided pattern searches using compilations of known yeast tRNAs and tRNA genes (10; Kleine and Feldmann, unpublished data), which were sufficient to identify all genes for tRNA species the structures of which had been solved experimentally and, additionally, to detect genes for novel tRNA species by homology. Performing a whole-genome analysis of *S.cerevisiae* strain αS288C by this approach we were able to identify the 274 intact nuclear encoded tRNA genes and an additional tRNA(Asp) gene on chromosome XIV interrupted by a Ty1 element near its 5′-end (1). An analysis based on the computer-aided program Pol3scan identified 274 (intact) tRNA genes (11), whereas application of previous tools tailored for identifying tRNA genes, such as tRNAscan (12), FAStRNA (13) and tRNAscan-SE (14), did not detect some tRNA genes exhibiting peculiar features. Our compilation of nuclear yeast tRNA genes and further detailed information are accessible at the MIPS Web site (http://www.mips.biochem.mpg.de/yeast/).

A catalogue of the 24 yeast mitochondrial tRNA genes is also presented at http://www.mips.biochem.mpg.de/yeast/ . Only 14 of the corresponding tRNA products have been sequenced. Nineteen of the mitochondrial tRNA genes could be identified by the tRNA scanning programs; prediction of the remaining five tRNA genes was not accomplished due to the fact that in many cases mitochondrial tRNA structures deviate from the general clover-leaf structure and that the genetic code in yeast mitochondria deviates from the universal code (see for example 15). For example, the gene for $tRNA^{Trp}$, which reads the codon UGA in yeast mitochondria, was erraneously predicted by tRNAscan-SE to be a $tRNA^{Sec}$ gene.

The locations of the nuclear tRNA genes on single chromosomes are shown in Figure 1 (for details see http://www.mips.biochem.mpg.de/yeast/). The tRNA genes are found scattered throughout the whole genome as single transcriptional units with the exception of four co-transcribed $tRNA^{Arg–Asp}$ 'pairs' (16). This finding and the actual number of tRNA genes agrees well with earlier experimental data (17). In contrast, tRNA genes in *Schizosaccharomyces pombe* may be organized in clusters (18). The average occurrence in the total genome is one tRNA gene in 45 kb, but, as expected, tRNA gene density varies for the single chromosomes (Fig. 1).

*To whom correspondence should be addressed. Tel: +49 89 5996 451; Fax: +49 89 5996 316; Email: feldmann@dec1.bio.med.uni-muenchen.de

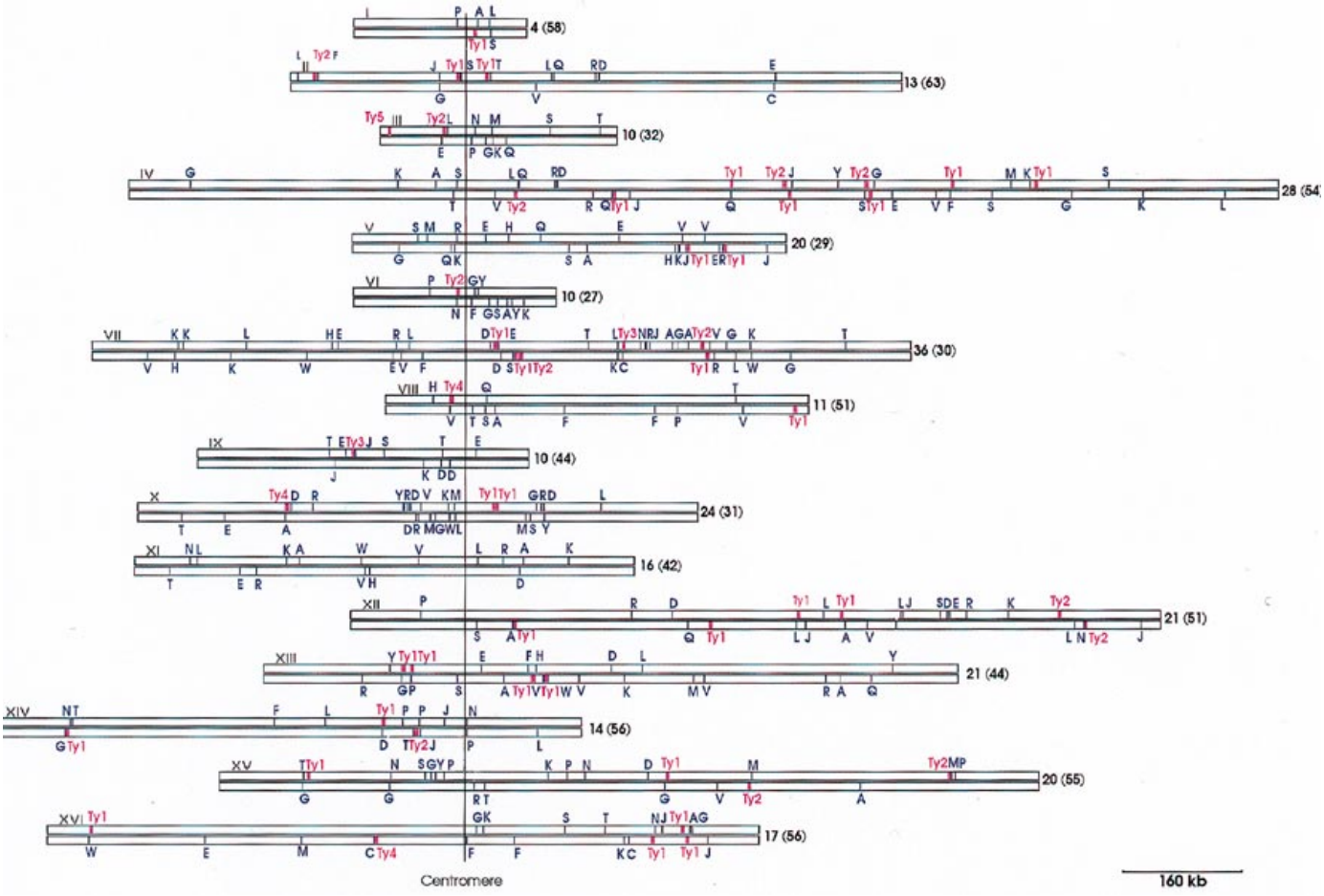

Figure 1. Location of tRNA genes and Ty elements on yeast chromosomes. The two open bars for each chromosome represent the Watson (upper) and Crick (lower) strands respectively. The tRNA genes are identified by the one letter code of the amino acid accepted by the corresponding tRNA (cf. Table 1). Ty1–Ty4 represent the different classes of retroelements. Transcription of the genetic entities shown on the upper strand is from left to right; those shown on the lower strand are transcribed from right to left. The numbers of tRNA genes for each chromosome are specified (average kb per tRNA gene in brackets).

tRNA families, tRNA gene redundancy and tRNA gene variants

By several criteria the 274 yeast tRNA genes can be considered as active genes and grouped into 42 families (Table 1) of distinct codon specificity. The two isoacceptors for methionine are assigned to distinct families, as the initiator $tRNA^{Meti}$ and the elongator $tRNA^{Met3}$ are clearly distinguished both by primary structure and function (cf. 11). In the majority of cases the single copies of a family share identical sequences within the tRNA structural part. Some families, however, comprise gene copies with slightly variant sequences giving rise to alternate base pairs in stems or alternate nucleotides at single sites of the tRNAs. Likewise, variations of the intron sequences can be observed in different members of a family. Altogether, sequence variants (including intron variants) exist in 11 tRNA gene families. The $tRNA^{Ser3}$ gene sequences represent special cases: in two of them the first C is missing in the anticodon loop (normally reading 5′-CTGCTAA-3′). If this ‘loss’ is not simply due to sequencing errors, the resulting tRNA should be a real variant that has only 6 nt in its anticodon loop. Genes for nonsense or missense suppressor tRNAs may be included into their ‘parent’ families, because they result from particular mutations in individual gene copies of this family. Though no suppressor tRNA genes are present in αS288C, we have included information on the provenance of such genes in suppressor strains in the two listings at http://www.mips.biochem.mpg.de/yeast/ , adopting data from the literature (19) and consulting the most recent genetic map (20).

Interestingly, we did not detect any tRNA gene the product of which would exhibit the unusual structural features typically found in the $tRNAs^{Sec}$ inserting selenocysteine in other eukaryotes (21,22). Furthermore, in strain αS288C no genes were detected for three (variant) tRNAs the primary structures of which have been reported in the literature: $tRNA^{His1}$, $tRNA^{Ser1}$ and $tRNA^{Thr1b}$. Two explanations are offered: (i) these tRNAs are encoded by additional genes which may be present in other yeast strains; (ii) these (and maybe other) variant tRNA structures are generated by RNA editing. The first possibility is supported by the fact that ‘soluble’ RNA from different strains had served as starting material in purifying specific tRNAs for sequencing. Moreover, there were numerous indications from tRNA profiles in different yeast strains (including haploid and diploid strains of different mating types) that the number and location of certain isoacceptors was variable. The second assumption is difficult to prove at present. To our knowledge editing of nuclear RNA in yeast has not been reported thus far.

Table 1. Yeast tRNA families and their genes

	tRNA species [a]	Number of functional genes [b]	tDNA anti-codon	tRNA anticodon [c]	Probable codon preferences [d]	Remarks on variant tRNA or tDNA sequences [e]
1	Ala1	11	AGC	IGC	GCU; GCC	
2	Ala2	5	TGC	*UGC	GCA; GCG?	
3	Arg1	1	CCT	CCU	AGG	
4	Arg2	6	ACG	ICG	CGU; CGC; CGA?	
5	Arg3	11	TCT	^{mem5}UCU	AGA	Four dimeric(Arg-Asp) genes; no gene for a variant tRNA(Arg3) in αS288C
6	Arg4	1	CCG	*CCG	CGG	
7	Asn	10	GTT	GUU	AAU; AAC	
8	Asp	15	GTC	GUC	GAU; GAC	Four dimeric(Arg-Asp) genes
9	Cys	4	GCA	GCA	UGU; UGC	
10	Gln1 Gln1a	7 2	TTG TTG	*UUG *UUG	CAA CAA	These variants to tRNA(Gln1) have 3 alternate bases in AC stem
11	Gln2	1	CTG	*CUG	CAG	Essential gene, closely related to Gln1.
12	Glu3	14	TTC	^{mom5s2}UUC	GAA	
13	Glu4	2	CTC	*CUC	GAG	
14	Gly1	16	GCC	GCC	GGU; GGC	
15	Gly2	3	TCC	NCC	GGA	
16	Gly3	2	CCC	*CCC	GGG	
17	His2	7	GTG	GUG	CAU; CAC	
18	Ile1	2 (i)	TAT	*UAU	AUA	Probably one variant gene
19	Ile2	13	AAT	IAU	AUU; AUC	
20	Leu1	3 (i)	TAG	UAG	CUA; CUG? CUU; CUC UUA; UUG	Variant intron sequences in one gene copy
21	Leu3	10(i)	CAA	^{ms}CAA	UUG	Variant introns in five gene copies
22	Leu4	7	TAA	NAA	UUA; UUG	
23	Leu5	1	GAG	*GAG	CUU; (CUC?)	
24	Lys1	14	CTT	CUU	AAG	Probably two variant genes
25	Lys2	7 (i)	TTT	^{ommm5}UmUU	AAA	Variant intron in one gene copy
26	Meti	5	CAT	CAU	AUG	
27	Met3	5	CAT	CAU	AUG	Variant tRNA(Met3) observed in other strain
28	Phe Phe1a	8 (i) 2 (i)	GAA GAA	G^{m}AA G^{m}AA	UUU; UUC UUU; UUC	Variant intron sequences tRNA(Phe1a) has alternate bases in acceptor stem; genes have variant introns
29	Pro1	10 (i)	TGG	$^{?}$UGG	CCA; CCG?	One variant gene; variant introns in 5 gene copies
30	Pro2	2	AGG	prob. IGG	CCU; CCC	
31	Ser2	11	AGA	IGA	UCU; UCC	
32	Ser3	4 (i)	GCT	*GCU	AGU; AGC	2 genes with variant AC loop (cf. Text)
33	Ser4	3	TGA	$^{?}$UGA	UCA	
34	Ser5	1 (i)	CGA	*CGA	UCG	
35	Thr1a	11	AGT	IGU	ACU; ACC	
36	Thr2	1	CGT	*CGU	ACG	
37	Thr3	4	TGT	*UGU	ACA	Probably one variant gene
38	Trp	6 (i)	CCA	C^{m}CA	UGG	
39	Tyr	8 (i)	GTA	GψA	UAU; UAC	Variant introns in 3 gene copies
40	Val1a Val1b	13 1	AAC AAC	IAC IAC	GUU; GUC GUU; GUC	Variant tRNA(Val1) gene
41	Val2a	2	TAC	^{ncm5}UAC	GUA	
42	Val2b	2	CAC	CAC	GUG	
	Total	**274**				

[a]As far as possible designations of the tRNA species follow those in the literature. For simplicity isoaccepting tRNA species that have not been sequenced earlier or were predicted from the gene sequences have been numbered arbitrarily.

[b](i) indicates the presence of intron sequences.

[c]Nomenclature for modified bases is as in Steinberg *et al.* (10). An asterisk indicates that this base has been deduced from the DNA sequence but that further information on whether this is a modified or unmodified base is lacking because the corresponding tRNA has not been sequenced.

[d]Exact *in vivo* codon recognitions have been determined experimentally in only a few cases. We largely follow the conventions proposed in Guthrie and Abelson (5). Question marks refer to the codons discussed in the text.

[e]Variants giving rise to suppressor tRNAs are not listed here.

Sequences and references for the tRNAs and tDNAs sequenced previously can be found at http://www.mips.biochem.mpg.de/yeast/

Introns and processing of precursor tRNAs

Intervening sequences in eukaryotic tRNA genes were first described for yeast tRNA^{Tyr} (23). With the availability of further sequences, rules for the location of introns within the coding sequences have been established, the mechanism of intron splicing in yeast has been elucidated and the components involved in this process have been characterized. Similarly, 5′- and 3′-end maturation and modification have been studied in great detail (for an overview see 24). The present catalogue reveals that 57 (21%) of the nuclear yeast tRNA genes contain sequences specifying introns in the tRNA precursors; in contrast, none of the mitochondrial tRNA genes contain an intron sequence. Surprisingly, introns are found in only 10 of the nuclear tRNA gene families (Table 1), but any one member of a particular family carries an intron. Generally the introns for a given tRNA species are identical or demonstrate very high sequence similarity. In a few cases, however, the introns are variable in length (± 1 nt) as well as in sequence (http://www.mips.biochem.mpg.de/yeast/).

It is still an unsolved question why introns are present in particular pre-tRNAs (24). Even with a knowledge of the complete catalogue of tRNA genes, no rules become obvious. As studied in certain suppressor tRNAs, some modification activities appear to require intron-containing tRNAs as substrates. In particular, introns may be required to introduce anticodon first base modifications during processing of particular pre-tRNAs such that the mature tRNAs are capable of reading selected codons. However, this requirement has been shown not to be universal and modifications to other positions in the tRNA sequence do not require precursor tRNAs as substrates, though many of the modifications normally occur at the level of precursors. On the other hand, the role of particular structural features of the introns for excision and ligation in the splicing process is well documented (24). An interesting observation was that introns can be shuffled among homologous tRNA genes (25). Little information is available on how intervening sequences might affect the level of tRNA gene expression.

tRNA multiplicity and codon capacity in yeast

Earlier studies have shown that yeast, like other organisms, employs nuclear encoded isoacceptors for the majority of the 20 amino acids, to meet the degeneracy of the genetic code. It became evident that ‘wobbling’ in codon–anticodon interactions allows a reduction in the number of tRNAs needed to account for the entire decoding capacity of an organism (26) and that ‘fine-tuning’ of codon–anticodon interactions are mediated by highly specific anticodon base modifications in particular tRNAs. Particular wobble rules for third position codon–anticodon pairing proposed for *S.cerevisiae* lead to the prediction of ~46 different tRNAs (5). The set of 42 tRNA families with distinct codon specificities now established for yeast is in good agreement with this prediction.

As can be seen from the currently available information summarized in Table 1, 57 of the 61 amino acid-specifying codons can be correlated with individual tRNAs. No particular isoacceptors have been found for four codons: GCG (Ala); CGA (Arg); CUG (Leu); CCG (Pro). Although these codons are very rare in yeast, they are present to a ‘normal’ extent in protein coding genes with low codon adaptation index (CAI) values (see below). Thus we have to assume that these triplets are decoded by the available isoacceptors. tRNA^{Ala2}, which probably has a modified U at the first anticodon position, should be capable of reading the codons GCA and GCG. For CGA (Arg) there are two alternatives: either tRNA^{Arg2} (anticodon ICG) is able to read three codons or, less likely, tRNA^{Arg4} (containing a C or modified

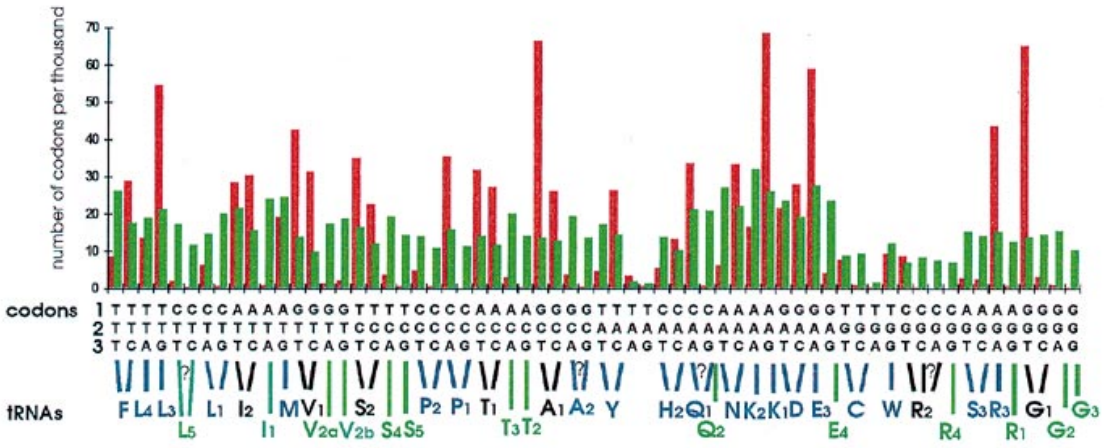

Figure 2. Codon usage in highly and lowly expressed yeast genes. The tRNAs reading particular codons are identified by the single letter code of the cognate amino acids accepted by them; suffix numbers are used to distinguish isoacceptors (cf. Table 1). Red bars, average of 263 highly expressed genes; green bars, average of 264 lowly expressed genes.

C at the first anticodon position) is able to decode CGA in addition to CGG. $tRNA^{Leu1}$ (unmodified U at the first anticodon position) is exceptional in that it has been shown to read all six synonomous leucine codons (27). $tRNA^{Pro1}$ contains an unknown modified U at the first anticodon position, which should be able to base pair with either A or G.

Correlation of tRNA abundance with gene copy number

Exact quantifications of all individual components of the tRNA population are not available. An estimate of the cellular content of 21 of 24 sequenced tRNAs dates back to 1982 (28) and does not include any one of the rare tRNAs. In a recent study Percudani *et al.* (11) found that the gene copy number for individual tRNA species correlates well with the previously measured intracellular content of the 21 tRNA species. This, together with the finding that there is a close link between tRNA gene redundancy and the overall amino acid composition of yeast proteins, indicates that intracellular levels of tRNAs in normally growing yeast cells are mainly determined by gene copy number (11).

However, variations in tRNA content in yeast at different growth rates has not been studied in detail, as has been done recently for example in *Escherichia coli* (29). Only for particular yeast tRNAs, such as the serine isoacceptors, have the relative abundances been found to vary in cells growing at different rates (30). Recent *in vitro* studies have suggested that both TFIIIB70 and TFIIIC play a role in coordinating the level of polymerase III transcription with cell growth rate (31).

It should also be noted that variations in copy number of major tRNA genes have been observed in different yeast strains but appear to be tolerated. Functional analysis of the five repeated methionine initiator tRNA genes (32) or the eight $tRNA^{Tyr}$ genes (33) have shown that knockout of up to two or five of the copies of these genes respectively is not growth limiting. These findings suggest that more intricate mechanisms must exist which allow the tRNA content to be adapted by regulating expression of the single tRNA gene copies. However, the problem of how expression of individual tRNA genes is regulated *in vivo* is still unsolved, because of the redundancy of tRNA genes.

tRNA gene redundancy and codon selection in yeast

Studies in the early 1980s pointed to a strong correlation between the abundance of yeast tRNAs and the occurrence of the respective codons in protein genes: genes which were known to be strongly expressed were found to be more biased than genes with a lower level of expression (28,34). These findings have been confirmed in a comprehensive study investigating the correlation of tRNA gene redundancy and translational selection in yeast by using a sample of 1756 distinct protein coding sequences: a significantly stronger co-adaptation between codon choice and tRNA gene copy number was observed in highly expressed genes (11).

By an analysis employing the CAI (35,36) we arrived at a similar conclusion. In brief, we have determined amino acid composition, number of individual codons and CAI for each protein in the total set. From this the average occurrence of codons for 263 entries with the highest CAI values (>0.5, group I) and the same number of entries with the lowest CAI values (<0.09, group II) respectively were calculated. Codon preferences in group I genes were found to be extremely biased, whereas codon usage in group II genes was nearly evenly distributed (Fig. 2; tables at http://www.mips.biochem.mpg.de/yeast/). In keeping with the general tendency of yeast to avoid G/C codons with either G or C in their second or third positions, these latter codons are virtually absent from group I protein genes, while they are used to a similar extent as their synonomous codons in group II genes. All group I genes are known to be highly expressed, whereas low expression levels have not been confirmed for all of the group II genes. Despite this uncertainty, it remains conclusive that the rare codons are more or less 'reserved' for particular yeast proteins. Obviously, these codons are decoded by tRNAs which are represented by only one or a low number of gene copies (Table 1).

Genomic organization of yeast tRNA genes and evolutionary aspects

With analysis of the complete genome gene redundancy in yeast has been established as a general phenomenon for the protein encoding genes (1,37). In most cases the duplicated sequences are

confined to the coding region of these genes and do not extend into the intergenic regions. Often the corresponding gene products share high similarity in terms of amino acid sequence or sometimes are even identical and, therefore, may be functionally related.

Thirty seven of the 42 yeast tRNA families are represented by multiple gene copies scattered throughout the genome. Alignments of the gene sequences encoding the isoaccepting tRNAs (http://www.mips.biochem.mpg.de/yeast/) reveals that the single copies share a particular feature with the protein encoding genes: in the majority of cases little if any homology is seen in the flanking regions between different members of a family nor is it seen between members from different gene families. At most some 10 or so base pairs are conserved within the immediate upstream and/or downstream regions of different members of a tRNA gene family. There are a few exceptions in which more extended similarities in the flanking regions are seen (e.g. for the two $tRNA^{Ile1}$ genes and for two gene copies of $tRNA^{Pro1}$ on chromosomes XII and XIII respectively), suggesting that dispersion of these gene copies by duplication events has occurred only recently.

In terms of basal tRNA gene transcription, which is governed by internal promoter sequences (38,39), sequence divergence in the flanking regions is not so surprising. However, considering the fact that transcription factor TFIIIB binds to sequences preceding the structural part of tRNA genes, 'fine-tuning' of tRNA gene expression should depend on the character of these sequences (see for example 40).

While maintenance of tRNA gene multiplicity is easily explained by interlocus gene conversion (41), questions as to the mechanism underlying establishment of tRNA gene multiplicity remain largely unsolved. A process mediated by reverse transcription seems plausible: the preservation of introns does not obviate this hypothesis, as splicing is a late step in tRNA maturation and precurors to tRNAs may be considered as substrates for retrotranscription (33). Other dispersal mechanisms, such as retrotransposon-mediated events, have been proposed to be involved in tRNA gene amplification (42), but our present knowledge of the organization of yeast tRNA genes renders this possibility unlikely.

Sequence comparisons of the entire chromosomes revealed the occurrence of so-called cluster homology regions (CHRs) in which homologous genes are arranged in the same order, with the same relative transcriptional orientation on two or more chromosomes (1,43). Wolfe and Shields (44) defined 55 such blocks and suggested a convincing model that explains the extended homologies seen in todays yeast by (allo- or auto-)tetraploidization between two ancient strains of *S.cerevisiae* followed by reciprocal translocations and 85% deletions of duplicated genes.

We reported earlier (1) that the largest CHRs share a number of tRNA gene 'pairs', both in the same relative location and orientation. The observed patterns imply that these particular tRNA genes have been duplicated together with neighbouring protein coding genes. Supposing that tRNA genes are an evolutionarily old aquisition and considering the fact that the flanking regions of these tRNA gene pairs have completely diverged, our finding agrees well with the notion that the CHRs were generated by ancient gene duplications. However, considering the fact that a large variety of duplicated gene copies, protein coding as well as tRNA genes, are found scattered throughout many locations in the yeast genome, one has to postulate additional mechanisms of gene amplification and dispersion. Particularly for the tRNA genes, the majority of which (232/274) are located outside CHRs, one would have to invoke two different types of mechanism by which duplications have been generated: (i) co-evolution of larger chromosomal segments containing both protein encoding and tRNA genes; (ii) insertion of 'duplicated' tRNA gene copies at singular sites in other chromosomes, a process which might have become operative after speciation.

Closer inspection of the seven largest CHRs revealed that they are composed of (short) 'units' each one containing one or two homologous protein encoding genes adjacent to a tRNA gene (examples are illustrated at http://www.mips.biochem.mpg.de/yeast/). This might suggest that such units were assembled prior to duplication of the entire blocks. One could even envisage the possibility that such units have contributed to duplications throughout evolution. Most strikingly, there is one example supporting such a regime. Chromosome XIV carries two copies of the $tRNA^{Ile2}$ gene in conjunction with two highly conserved protein encoding genes, residing in two inverted 4.2 kb segments (positions 569831–574066 and 598156–602399 respectively) separated by 24 kb of non-repetitious sequence (http://www.mips.biochem.mpg.de/yeast/). This duplication, in evolutionary terms, must have occurred only recently, as these segments share 99% sequence identity. Since a large part of chromosome XIV, including one of the $tRNA^{Ile2}$ gene copies, is duplicated on chromosome IX (45), we checked whether the neighbourhood adjacent to this tRNA gene on chromosome IX has similarity to the 4.2 kb region on chromosome XIV. This was not the case, suggesting that this $tRNA^{Ile2}$ gene copy has been amplified independently.

YEAST RETROTRANSPOSONS

Classes of Ty elements

The genome of yeast strain αS288C contains 51 'canonical' retrotransposons falling into the four distinct classes Ty1–Ty4 (Table 2). The Ty5 element identified on chromosome III and the seven Ty5 remnants found in αS288C have to be considered copies no longer capable of transposition. In contrast, transposition-competent Ty5 elements have been found in *S.paradoxus* (46,47).

Altogether the Ty elements occupy some 300 kb of genomic sequence in *S.cerevisiae*, equivalent to 2.4% of the genome. Additionally, 268 'solo' elements (i.e. LTRs of the various Ty elements) or remnants thereof have been identified in αS288C, marking former transposition/excision events. As seen from alignments of their DNA and protein sequences (http://www.mips.biochem.mpg.de/yeast/), the Ty4 elements share considerable sequence identity, whereas the Ty1 and the Ty3 (48) elements appear to fall into two subclasses each. Five of the 33 Ty1 copies appear to be non-intact elements (Table 2). In seven of the Ty1 copies deviations from the consensus sequences were encountered, which may result from sequencing errors. The same is true for one of the three Ty4 copies. If the sequences are corrected for these ambiguities, translation would result in correct TYA/TYB proteins.

Genomic organization and transposition activity of retroelements

Early data already pointed to a close association between Ty elements and tRNA genes (49,50). These notions were substantiated by an analysis of a large variety of tRNA gene loci showing that the 5′-flanking regions of tRNA genes were the preferred target sites for Ty transposition (42,50–53). In several cases complex patterns of Ty elements and remnants thereof were

Table 2. Retrotransposons in the yeast genome

Chromosome	Ty1	Ty2	Ty3	Ty4	Ty5	Total
I	A[a] (1)					1
II	BL[v], BR[a] (2)	B (1)				3
III		C[a] (1)			C (1)	2
IV	DR1, DR2[bv], DR3, DR4[b], DR5, DR6 (6)	DR1,DR2,DR3 (3)				9
V	ER1, ER2[b] (2)					2
VI		F (1)				1
VII	GR1, GR2, GR3[a] (3)	GR1[a],GR2 (2)	G[v] (1)			6
VIII	H[i] (1)			H (1)		2
IX			I (1)			1
X	JR1, JR2 (2)			J[a] (1)		3
XI						0
XII	LR1, LR2, LR3, LR4[a] (4)	LR1,LR2[ab] (2)				6
XIII	ML1[b], ML2, MR1[v], MR2[b] (4)					4
XIV	NL1[a], NL2[v] (2)	N (1)				3
XV	OL, OR (2)	OR1,OR2 (2)				4
XVI	PL; PR1, PR2[a], PR3 (4)			P[a] (1)		5
total found	**33**	**13**	**2**	**3**		**52**
in 'old'site	15	11	1	3		31
in 'new'site	18	2	1	0		21
subtypes	two [v])	one	two	one		
solo elements or remnants						268

Nomenclature of the Ty elements is as in the sequence annotations by MIPS (1) (e.g. ML1 is the first element on the left arm of chromosome XIII); number of elements on a given chromosome are in brackets.
[a]Sequence ambiguities compared with consensus; probably sequencing errors.
[b]Non-intact element.
[i]Insert of 114 bp corresponding to 38 amino acids (cloning artefact?).
[v]Ty1_BL, Ty1_DR2, Ty1_MR1 and Ty1_NL2 belong to a second subtype of Ty1 elements having a variant TYA protein sequence.

encountered, representing the footsteps of multiple transposition and excision events at certain 'hotspots' of transposition. While Ty1, Ty2 and Ty4 integrate in a region-specific manner, Ty3 integration occurs in a distance-specific manner, 16–19 bp upstream of tRNA genes (50,54). Experimental proof for tRNA gene regions in yeast (55) or upstream sequences of polymerase III transcribed genes in general (9) as the preferred target sites for Ty1 transposition has been provided. Ty3 is invariably targeted 1–4 bp from the initiation site of polymerase III transcription (56).

Our analysis revealed that 26% (71/274) of the tRNA genes are found closely associated with one single element, while 34% of the tRNA genes (92/274) are located in 'hotspots of transposition' in which up to 10 elements accumulated in complex patterns. While all Ty4 and Ty3 elements are located in the upstream flanking regions of tRNA genes, eight of the full-length Ty1 and Ty2 elements are located outside tRNA flanking regions. One of these elements became integrated next to a 5S RNA gene and another disrupts an open reading frame (YJL181w) on chromosome X. The distribution of Ty elements in the genome of αS288C appears not to be completely random, in that half (27/51) are located less than 100 kb from centromeres or telomeres. Transpositional activity of the different classes of the Ty elements should be reflected in the number of *de novo* integration sites. No such sites are found for Ty4, in accordance with the observation that this element reveals very low transcriptional activity and hence little propensity for transposition (57). *De novo* integration next to a tRNA gene occurred for one of the two Ty3 elements. The majority of the Ty2 elements (11/13) are maintained in 'old' sites; in only three cases has integration occurred at 'new' sites. For Ty1 about half (18/33) of the elements reside in 'pre-existing' sites, while the rest have been targeted to 'virgin' sites (the majority of these representing tRNA gene flanking regions). These data clearly demonstrate the high transpositional capacity of Ty1.

Ty expression and codon usage in Ty open reading frames

Recent work has established that tRNA genes and tRNAs influence several key steps in transposition (for an overview see 7). Like many retroviruses, elements Ty1–Ty4 employ translational frameshifting and rare tRNAs to regulate expression of their *gag* (TYA) and *pol* (TYB) gene products. Specific tRNAs are also needed as primers for reverse transcription. In Ty1, Ty2 and Ty3 the primer binding site is complementary to 13 nt of the acceptor stem of the initiator $tRNA^{Met}$, a feature shared by most of the members of the Ty1/*copia* group of retrotransposons. An exception to this rule appears to be the Ty4 element, which has a primer binding site complementary to 18 nt of the acceptor stem of $tRNA^{Asn}$.

Codon usage in Ty expression is similar to that of the average of all yeast proteins (http://www.mips.biochem.mpg.de/yeast/), which means that the Ty elements fall into the category of intermediately expressed genes. However, for some codons there are substantial deviations from the average codon preference in yeast genes. This might suggest that codon usage in the Ty elements is not fully adapted to that of their host. We have also noticed in an analysis using sliding windows (Feldmann, unpublished data) that the average G+C content of particular portions of genes encoding the TYB proteins considerably differ from the average G+C content of host genes. As in highly expressed yeast genes, there is a tendency in one or other element to avoid particular G/C-rich codons with a 3′ C or G.

Interactions between retroelements and the host

A strict balance is needed between level of transposition to maintain a population of transposition-competent elements and the amount of activity that can be tolerated by the organism. Obviously, in yeast such a compromise has been reached by a mechanism that directs integration of the elements to target sites which are least hazardous for the yeast cell, i.e. the upstream flanking regions of tRNA genes. Whole-genome analysis reveals that these regions are normally extended and devoid of protein coding genes. While the target site specificities in transposition of all types of Ty elements have been clearly established by sequence analysis and experimentation, mechanisms that dictate integration specificity are less well understood. In this regard the best studied example is that of Ty3, for which target preference has been shown to be dependent upon functional promoter elements within the polymerase III transcribed gene (56,58,59). Recent findings point to the possibility that retroelement integration is generally also affected by chromatin structure and by interactions ('tethering') between the integrase and chromatin-associated proteins (for a review see 8). The *Saccharomyces* Ty5 retrotransposon family was found to exhibit a target preference different from that of the other Ty elements: Ty5 target specificity seems attributable to interactions between transposition intermediates and constituents of silent chromatin, as integrated elements are largely found adjacent to transcriptional silencers or type X telomeric repeats (46,47).

The strong target selection for integration of Ty elements raises the question whether there is more to this preference than physical interactions. Given the fact that upstream flanking regions are involved in modulating tRNA gene expression, the problem has

been addressing whether Ty elements or their relics could participate in this type of modulation. It has been shown (60) that pre-tRNA levels were moderately increased when the target gene was associated with a Ty3 or σ element. We have shown that the transcriptional activity of several tRNA genes was dependent on the type of Ty1 element associated with them (61–63). We could further demonstrate (64; Krieg and Feldmann, unpublished data) that the tRNA transcription complex acts as a boundary in positioning of upstream nucleosomes and that efficient tRNA gene transcription correlates with an array of positioned nucleosomes. Many different upstream sequences allow formation of this nucleosomal organization, but Ty1 LTRs turned out to be the most efficient. Our results also support the view that modulatory effects *in vivo* do not involve specific upstream sequence elements, but that the particular chromatin structure formed over the extended tRNA gene region plays the dominant role. Similarly, Morse *et al.* (65) have shown that tRNA genes fused to either of two nucleosome positioning signals (such that the predicted nucleosome would either incorporate near its centre or to the tRNA start site) and introduced into yeast cells remained competent only when not incorporated into positioned nucleosomes.

Recent work has revealed yet another aspect of interaction between retroelements and the host, namely involvement of Ty1 in chromosome healing, which occurrs by repair of double-strand lesions by non-homologous recombination (66,67). In all, the non-random integration of retrotransposons suggests that genome organization, probably chromatin structure, influences target choice. In turn, targeted integration should affect genome organization. Ultimately this interdependence strengthens the impression that retroelements and the host have developed symbiotic behaviours.

CONCLUSIONS AND PERSPECTIVES

The availability of the entire sequence of the *S.cerevisiae* genome provides detailed insight into the genomic distribution and organization of tRNA genes and Ty elements. This information illustrates the intricate interactions between Ty elements and tRNA genes and contributes to a better understanding of the correlation between tRNA gene redundancy and translational selection in yeast. The rules found for codon adaptation in yeast will be useful to exploit similar correlations in other eukaryotic systems. As we experienced, however, sequence information by itself is insufficient to provide any clue to a variety of interesting problems. For example, the maintenance and significance of introns in particular tRNAs or mechanisms that are involved in amplification and dispersal of tRNA genes are extremely difficult issues. Now that we know the number of tRNA genes and their molecular environments, experimental approaches might be designed to gain a better insight into the correlation between tRNA gene expression and tRNA content. Generation of null mutants for rare tRNA genes will probably reveal whether the corresponding tRNAs are absolutely required for yeast decoding capacity or whether they can be substituted by appropriate isoacceptors. Finally, yeast genetics is required to reveal all the intricacies of the interplay between tRNAs and Ty retrotransposons.

ACKNOWLEDGEMENTS

Work on the yeast genome project was supported by the European Commission and, for the German participants, by the Bundesminister für Forschung und Technologie. Work on bioinformatics for this project was supported by the European Commission, the Max-Planck-Gesellschaft, the Forschungszentrum für Umwelt und Gesundheit and the Yeast Industry Platform. J.H. was supported by a grant from the European Community (BIO4-CT95-0080). We would like to thank H.W.Mewes for discussions and K.Kleine for initial work on the compilation of yeast tRNA genes.

REFERENCES

1 Goffeau,A. *et al.* (1996) *Science*, **274**, 546–567.
2 Goffeau,A. *et al.* (1997) *Nature*, **387** (suppl.), 1–105.
3 Clayton,R.A., White,O., Ketchum,K.A. and Venter,J.C. (1997) *Nature*, **387**, 459–462.
4 Oliver,S.G. (1996) *Trends Genet.*, **12**, 241–242.
5 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 Press, Cold Spring Harbor, NY, pp. 487–528.
6 Boeke,J.D. and Sandmeyer,S.B. (1991) In Broach,J.R., Pringle,J. and Jones,E.W. (eds), *The Molecular and Cellular Biology of the Yeast Saccharomyces cerevisiae*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 193–261.
7 Voytas,D.F. and Boeke,J.D. (1993) *Trends Genet.*, **9**, 421–427.
8 Curcio,M.J. and Morse,R.H. (1996) *Trends Genet.*, **12**, 436–438.
9 Devine,S.E. and Boeke,J.D. (1996) *Genes Dev.*, **10**, 620–63.
10 Steinberg,S. Misch,A. and Sprinzl,M. (1993) *Nucleic Acids Res.*, **21**, 3011–3015.
11 Percudani,R., Pavesi,A. and Ottonello,S. (1997) *J. Mol. Biol.*, **268**, 322–330.
12 Fishant,G.A. and Burks,C. (1991) *J. Mol. Biol.*, **220**, 659–671.
13 El-Mabrouk,N. and Lisacek,F. (1996) *J. Mol. Biol.*, **264**, 46–55.
14 Lowe,T.M. and Eddy,S.R. (1997) *Nucleic Acids Res.*, **25**, 955–964.
15 Ueda,T. and Watanabe,K. (1993) *Origin Life Evol. Biosphere*, **23**, 345–364.
16 Schmidt,O., Mao,J., Ogden,R., Beckmann,J., Sakano,H., Abelson,J. and Söll,D. (1980) *Nature*, **287**, 750–752.
17 Feldmann,H.(1977) *Nucleic Acids Res.*, **4,** 2831–2841.
18 Kuhn,R. M., Clarke,L. and Carbon,J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1306–1310.
19 Hinnebusch,A.G. and Liebman,S.W. (1991) In Broach,J.R., Pringle,J. and Jones,E.W. (eds), *The Molecular and Cellular Biology of the Yeast Saccharomyces cerevisiae*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 1, pp. 627–736.
20 Cherry,J.M. *et al.* (1997) *Nature*, **387** (suppl.), 67–73.
21 Schuster,C., Myslinski,E., Krol,A. and Carbon,P. (1995) *EMBO J.*, **14**, 3777–3787.
22 Sturchler-Pierrat,C., Hubert,N., Totsuka,T., Mizutani,T., Carbon,P. and Krol,A. (1995) *J. Biol. Chem.*, **270**, 18570–18574.
23 Goodman,H.M., Olson,M.V. and Hall,B.D. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5453–5457.
24 Hopper,A.K. and Martin,N.C. (1992) In Jones,E.W., Pringle,J. and Broach,J.R. (eds), *The Molecular and Cellular Biology of the Yeast Saccharomyces cerevisiae: Gene Expression.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 2, pp. 99–141.
25 Olson,M.V., Page,G.S., Sentenac,A., Piper,P.W., Worthington,M. Weiss,R.B. and Hall,B.D. (1981) *Nature*, **291**, 464–469.
26 Crick,F.H.C. (1966) *Cold Spring Harbor Symp. Quant. Biol.*, **31**, 3–10.
27 Weissenbach,J. and Dirheimer,G. (1977) *FEBS Lett.*, **82**, 71–76.
28 Ikemura,T. (1982) *J. Mol. Biol.*, **158**, 573–597.
29 Dong,H., Nilsson,L. and Kurland,C.G. (1996) *J. Mol. Biol.*, **260**, 649–663.
30 Heyman,T., Agoutin,B., Fix,C., Dirheimer,G. and Keith,G. (1994) *FEBS Lett.*, **347**, 143–146.
31 Sethy,I., Moir,R.D., Librizzi,M. and Willis,I.M. (1995) *J. Biol. Chem.*, **270**, 28463–28470.
32 Byström,A.S. and Fink,G.R. (1989) *Mol. Gen. Genet.*, **216**, 276–286.
33 Olson,M.V. (1991) In Broach,J.R., Pringle,J. and Jones,E.W. (eds), *The Molecular and Cellular Biology of the Yeast Saccharomyces cerevisiae*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 1, pp. 1–40.
34 Bennetzen,J.L. and Hall,B.D. (1982) *J. Biol. Chem.*, **257**, 3026–3031.
35 Sharp,P.M. and Li,W.H. (1987) *Nucleic Acids Res.*, **15**, 1281–1295.
36 Lloyd,A.T. and Sharp,P.M. (1992) *J. Hered.*, **83,** 239–240.
37 Dujon,B. (1996) *Trends Genet.*, **12**, 263–270.

38 Geiduschek,E.P. and Tocchini-Valentini,G.P. (1988) *Annu. Rev. Biochem.*, **57**, 873–914.
39 Hernandez,N. (1993) *Genes Dev.*, **7**, 1291–1308.
40 Kassavetis,G.A., Braun,B.R., Nguyen,L.H. and Geiduschek,E.P. (1990) *Cell*, **60**, 235–245.
41 Amstutz,H., Munz,W.-D., Heyer,U. Leupold,U. and Kohli,J. (1985) *Cell*, **40**, 879–886.
42 Warmington,J.R., Green,R.P., Newlon,C.S. and Oliver,S.G. (1987) *Nucleic Acids Res.*, **15**, 8963–8982.
43 Mewes,H.W. *et al.* (1997) *Nature*, **387** (suppl.), 9–32.
44 Wolfe,K. and Shields,D.C. (1997) *Nature*, **387**, 708–713.
45 Philippsen,P. *et al.* (1997) *Nature*, **387** (suppl.), 93–97.
46 Zou,S., Wright,D.A. and Voytas,D.F. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 920–924.
47 Zou,S., Ke,N., Kim,S.J. and Voytas,D.F. (1996) *Genes Dev.*, **10,** 634–645.
48 Hansen,I.J., Chalker,D.L. and Sandmeyer,S.B. (1988) *Mol. Cell. Biol.*, **8**, 5245–5256.
49 Eigel,A. and Feldmann,H. (1982) *EMBO J.*, **1,** 1245–1250.
50 Sandmeyer,S.B. and Olson,M.V. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7674–7678.
51 Warmington,J.R., Anwar,R., Newlon,C.S., Waring,R.B., Davies,R.W., Indge,K.J. and Oliver,S.G. (1986) *Nucleic Acids Res.*, **14**, 3475–3485.
52 Hauber,J., Stucka,R., Krieg,R. and Feldmann,H. (1988) *Nucleic Acids Res.*, **16,** 10623–10634.
53 Lochmüller,H., Stucka,R. and Feldmann,H. (1989) *Curr. Genet.*, **16**, 247–252.
54 Brodeur,G.M., Sandmeyer,S.B. and Olson,M.V. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3292–3296.
55 Ji,H., Moore,D.P., Blomberg,M.A., Braitermann,L.T., Voytas,D.F., Natsoulis,G. and Boeke,J.D. (1993) *Cell*, **73**, 1007–1013.
56 Chalker,D.L. and Sandmeyer,S.B. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 4927–4931.
57 Hug,A. and Feldmann,H. (1996) *Nucleic Acids Res.*, **274**, 2338–2346.
58 Chalker,D.L. and Sandmeyer,S.B. (1992) *Genes Dev.*, **6**, 117–128.
59 Kirchner,J. Connolly,C.M. and Sandmeyer,S.B. (1995) *Science*, **267**, 1488–1491.
60 Kinsey,P.T. and Sandmeyer,S.B. (1991) *Nucleic Acids Res.*, **19**, 1317–1324.
61 Nelböck,P., Stucka,R. and Feldmann,H. (1985) *Biol. Chem. Hoppe-Seyler*, **366**, 10410–1051.
62 Hauber,J., Nelböck,P., Pilz,U. and Feldmann,H. (1986) *Biol. Chem. Hoppe. Seyler*, **367**, 1141–1146.
63 Krieg,R., Stucka,R., Clark,S. and Feldmann,H. (1991) *Nucleic Acids Res.*, **19**, 3849–3855.
64 Feldmann,H. (1988) In Grunberg-Manago,M., Clark,B.F.C. and Zachau,H.G. (eds), *Evolutionary Tinkering in Gene Expression*, NATO ASI Series A. Plenum Press, NY, Vol. 169, pp.79–86.
65 Morse,R.H., Roth,S.Y. and Simpson,R.T. (1992) *Mol. Cell. Biol.*, **12**, 4015–4025.
66 Teng,S.-C., Kim,B. and Gabriel,A. (1996) *Nature*, **383**, 641–644.
67 Moore,J.K. and Haber,J.E. (1996) *Nature*, **383**, 644–646.