5 Yeast Genetic Structures and Functions

5.1 Yeast Chromosome Structure and Function

5.1.1 Yeast Chromatin

Compacting the genome. When biologists worked with bacteria it was evident to them that prokaryotes carry their genetic information on a single DNA molecule that is usually circular, and that this DNA is associated with (basic) proteins that condense and package the DNA. In eukaryotes, in which the genetic material is encapsulated in the cell nucleus and divided between a set of different chromosomes, the DNA had been found to be associated with other components that were amenable to package the DNA. However, for a long time it remained a mystery how the DNA was packaged in this structure collectively called **chromatin**. It was only in 1974 that the discovery of the **nucleosome** as a fundamental unit for chromatin organization (Kornberg and Lorch, 1999) began to establish our solid knowledge on chromatin structure, and how its organization contributes to chromosome replication and gene expression, both of which afford the transient availability of an "open chromatin configuration."

5.1.1.1 Organization of Chromatin Structure

How DNA goes with protein. As we have seen in Chapter 2, the nucleus of *Saccharomyces cerevisiae* accommodates **16 chromosomes**, each of which carries a centromere and two telomeres. The repeat unit of chromatin is the core nucleosome, which in yeast contains 146 bp of DNA wrapped around the histone octamer that consists of two molecules each of the core **histones** H2A, H2B, H3, and H4 (nearly identical genes for the histones occur in duplicate: *HTA1/HTA2*, *HTB1/HTB2*, *HHT1/HHT2*, and *HTF1/HTF2*, respectively). As in higher eukaryotes, nucleosomal arrays along the DNA fold into a 30-nm fiber; a single linker histone H1 gene, HHO1, has been found in yeast. In addition to the canonical histones, histone variants exist that are structurally related to the normal histones, but are functionally distinct, such as the centromeric nucleosome histone variant Cse4p, which is required for proper kinetochore function (*cf.* Section 7.1), and a **variant** of histone H2A, H2A.Z (gene *HTZ1*), exchanged in 5–10% of the nucleosomes by the SWR1 complex that is involved in transcriptional regulation through prevention of the spread of silent heterochromatin.

The distances between nucleosomes are not constant and may vary. Furthermore, there are heterochromatic regions that suppress transcription from resident genes. In such regions, additional proteins bind to the nucleosomes, which leads to gene silencing. In *S. cerevisiae*, the silent mating-type loci and the subtelomeric regions belong to this category; subtelomeric regions are found to be repressed by the presence of the SIR (silent information regulator) proteins and Rap1p.

Generally, the organization of DNA into chromatin and chromosomal structures plays a central role in many aspects of yeast cell biology – accessibility of the genetic material during replication, chromosome stability and segregation, gene expression, recombination, and DNA repair are intimately linked to chromatin configuration. Chromosome compaction is changed at mitosis (or meiosis), when cohesin and condensin proteins bind to chromatin, thereby inducing a more condensed status of chromosomes during cell division.

We have to recollect that in all cases the cellular machineries work on chromatin as the "native" DNA template. As chromatin is normally repressive to extraneous access, this inhibitory effect has to be surmounted by regulatory factors. However, the original chromatin structures have to be reinstalled as soon as possible after exertion of any of the above processes. Many of the mechanistic rules operating in the interaction between chromatin and modulating factors have been derived from the *S. cerevisiae* model.

5.1.1.2 Modification of Chromatin Structure

How to access the vault? Two main principles can be distinguished that regulate chromatin access: modification of the histones by various enzymatic activities (**chromatin-modifying complexes**) and temporary reorganization of the local nucleosome structure by **chromatin-remodeling complexes** (Figure 5.1).

5.1.1.2.1 Modification of Histones The histones are basic proteins that have a well-defined core domain and unstructured tail domains at both the N- and C-terminus. Although modifications may occur within the core domain, the majority of modifications concern amino acid residues in the tail regions.

Yeast: Molecular and Cell Biology, Second Edition. Edited by Horst Feldmann.

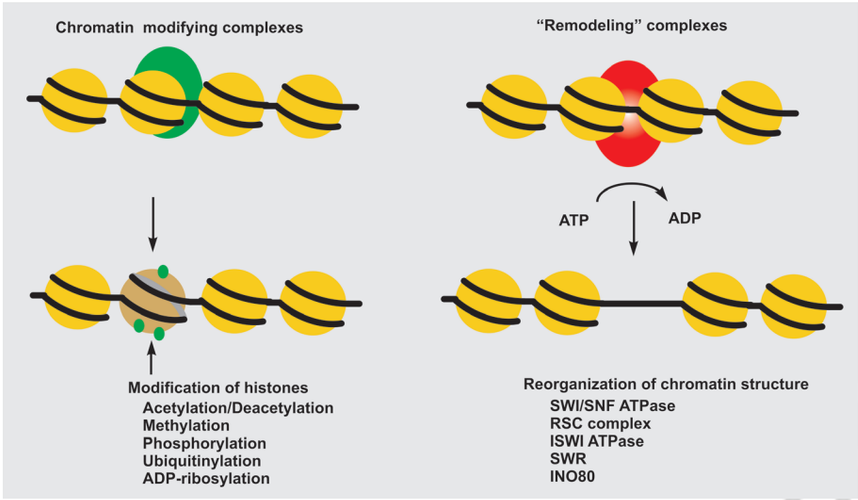


Fig. 5.1 Two modes of chromatin modulation.

Histone Acetylation

Masking lysine residues. Histone modifications can occur post-translationally at many sites along these basic proteins; preferred targets are the ε-NH_3^+ groups of lysine residues. Acetylations were among the first modifications that became recognized, because these modifications reduce the number of positive charges in histones. **Histone acetylation** works with acetyl-CoA as a donor coenzyme. Histone acetylation remains one of the best-studied post-translational modifications that have a role in many cellular processes.

Histone acetyltransferases (HATs) have been isolated in organisms ranging from yeast to humans. On the basis of sequence homology, each HAT falls into one of three categories (Table 5.1):

i) The Gcn5-related *N*-acetyltransferase (**GNAT**, general control nonderepressible 5-related acetyltransferase) family. In yeast, the GNAT family includes the Rtt109, SAGA, SLIK, ADA, and HAT-A2 complexes.

Table 5.1 Histone acetyltransferase complexes (HAT).

Family	Complex	Subunits	Functions
GNAT family	Rtt109		HAT that associates with transcriptionally active genes and is required for proper acetylation of H3K56, which occurs during both the premeiotic and mitotic S phase, and persists throughout DNA damage repair; stimulated by histone chaperone Asf1p, which governs the substrate specificity of Rtt109p
	SAGA	Gcn5p (Kat2)	HAT catalytic subunit
		Ada1p (Hfi1p)	adaptor protein for complex
		Ada2p	transcriptional coactivator
		Ada3p (Ngg1p)	expands range of lysines undergoing acetylation
		Ada4p	subunit
		Ada5p	subunit
		Spt3p	interacts with TATA-binding protein (TBP)
		Spt7p	essential subunit
		Spt8p	Controls TBP–TATA interaction at the promoter; positive and negative functions of Spt8p in transcription are mediated through the general transcription factor TFIIA
		Spt20p (Ada5p)	responsible for integrity of complex
		Tra1p	interacts with acidic activators
		Taf5p	subunit TFIID, polymerase II initiation
		Taf6p	subunit TFIID, similar to histone H4
		Taf9p	subunit TFIID, similar to histone H3
		Taf10p	subunit TFIID, polymerase II initiation
		Taf12p	subunit TFIID, polymerase II initiation, similar H2A
		Ubp8p	ubiquitin-specific protease for H2B deubiquitylation

Table 5.1 (*Continued*)

Family	Complex	Subunits	Functions
		Sgf11p	required for Ubp8p association with complex
		Sgf29p	potential subunit
		Sgf73p	potential subunit
		Sus1p	involved in mRNA export coupled transcription activation and elongation; subunit of TREX-2
		Rtg2p	sensor of mitochondrial dysfunction
		Chd1p	nucleosome remodeling factor; contains chromodomain and helicase domain
	SLIK	SAGA-like	lacks Spt8p and has truncated Spt7p
	ADA	Gcn5p	HAT catalytic subunit
		Ada2p	Ada2p increases HAT activity of Gcn5p
		Ada3p	expands range of lysines acetylated
		Ahc1p	subunit required for complex integrity
	HAT 2A	Hat1p	catalytic subunit; acts with acetyl-CoA
		Hat2p	effects high-affinity binding to free H4
MYST family	SAS	Sas2p	HAT catalytic subunit in antisilencing
		Sas4p	regulates transcriptional silencing; required for Sas2p activity
		Sas5p	regulates transcriptional silencing; stimulates Sas2p activity
	NuA3 (HAT complex that acetylates histone H3)	Sas3p	HAT catalytic subunit
		Nto1p	contains PHD finger domain that interacts with methylated histone H3
		Eaf6p	also subunit of NuA4
		Yng1p	contains PHD finger domain that interacts with methylated histone H3
		Taf14p	subunit of several complexes; required for efficient transcription and active in many regulatory complexes
	NuA4 (essential histone H4/H2A acetyltransferase complex)	Esa1p	catalytic subunit; required for cell cycle progression and transcriptional silencing at the rDNA locus
		Eaf1p	assembly platform; required for initiation of premeiotic DNA replication (Ime1p)
		Eaf3p	Esa1p-associated factor
		Eaf5p	Esa1p-associated factor
		Eaf6p	also subunit of NuA3
		Eaf7p	subunit
		Epl1p	component
		Tra1p	interacts with acidic activators
		Yaf9p	also subunit of SWR1 complex; interacts with Swc4; antagonizes silencing at telomeres
		Swc4p	component of the Swr1p complex that incorporates Htz1p into chromatin
		Yng2p	similar to human tumor suppressor ING1
		Arp4p	actin-like protein
		Act1p	actin
	Piccolo	Esa1p	as above in NuA4
		Epl1p	
		Yng2p	
Others		Nut1p	component of Mediator complex
		TAF1 (TAF250)	TFDII subunit with HAT activity
	Elongator complex of polymerase II	Elp1p (Iki3p)	major HAT component of RNA polymerase II holoenzyme responsible for transcriptional elongation; predominant acetylation sites H3K14 and H4K8; involved in chromatin remodeling; required for modification of wobble nucleosides in tRNA
		Elp2p	Elp3p is the catalytic subunit
		Elp3p	
		Elp4p	
		Elp5p (Iki1p)	
		Elp6p (Hap3p)	

GNAT, Gcn5-related acetyltransferase; MYST, yeast and human founding members MOZ, YBF2/SAS3, SAS2, and TIP60; SAGA, Spt–Ada–Gcn5 acetyltransferase; SLIK, SAGA-like; PHD, plant homeodomain.

ii) The second group of yeast HATs, **MYST**, derives its name from the yeast and human founding members MOZ, YBF2/SAS3, SAS2, and TIP60.

iii) There is a third family of **HATs** – the cytoplasmic Hat1p and the elongator component Elp3p in yeast, as well as the general transcription factor HATs including the TFIID subunit TAF1 in yeast (TAF250 in mammals), TFIIIC (the general transcription factor in the RNA polymerase III basal machinery), and a component of the Mediator complex, Med5p/Nut1p, in yeast; in mammals, p300/CBP, AFT-2, and so on, belong to this family.

Family connections. GNAT family members consist of HATs that have sequence and structural similarity to Gcn5p, and regulate the recruitment of transcription factors to their target promoters (Vetting *et al.*, 2005). In contrast, **MYST family** members are involved in the regulation of a variety of DNA-mediated reactions, such as promoter-driven transcriptional regulation (Utley and Cote, 2003), long-range/chromosome-wide gene regulation (Ehrenhofer-Murray *et al.*, 1997), DNA double-strand break (DSB) repair (Van Attikum and Gasser, 2005), and licensing of DNA replication (Iizuka *et al.*, 2006).

Most similar to *S. cerevisiae* Gcn5p are the cytoplasmic Hat1p and the elongator component Elp3p in yeast. Gcn5p is also the catalytic subunit of the **SAGA** transcriptional activation complex (Daniel and Grant, 2007). Sas3p is part of the **NuA3** (nucleosome acetyltransferase of H3) complex as the acetyltransferase subunit and the MYST HAT Esa1p, the only essential HAT in yeast, is the catalytic subunit of the **NuA4** (nucleosome acetyltransferase of H4) complex (Doyon and Cote, 2004; Doyon *et al.*, 2004). Yeast NuA4 is a 13-subunit HAT complex responsible for acetylation of histone H4 and H2A N-terminal tails (Figure 5.2). NuA4 can be recruited by activators *in vitro* and *in vivo* to create a large domain of histone H4/H2A hyperacetylation and activate transcription. The yeast NuA4 subunits (Tra1p, Eaf1p, Epl1p, Esa1p, Eaf2p, Yng2p, Arp4p, Act1p, Eaf3p, Yal9p, and Eaf5p–Eaf7p) have clear mammalian homologs, demonstrating the existence of a NuA4 complex in mammals (Doyon and Cote, 2004). Part of the yeast NuA4 subunits are also found in the yeast remodeling complex SWR1; in mammals, the NuA4 complex appears to have evolved from subunits of both complexes. Both Esa1p and its human counterpart, Tip60, have been linked to transcription regulation, as well as DNA DSB repair. Interestingly, several HAT complexes contain actin and actin-related proteins (Arps) – a notion that has lead to the hypothesis that they are directed at their site of action by the nuclear scaffold.

Rtt109p is a HAT that associates with transcriptionally active genes and is required for proper acetylation of histone H3 at K56, which occurs during both the mitotic and premeiotic S phase, and persists throughout DNA damage repair (Driscoll, Hudson, and Jackson, 2007; Fillingham *et al.*, 2008; Han *et al.*, 2007). This reaction is stimulated by histone chaperone Asf1p, which governs the substrate specificity of Rtt109p. Acetylation of H3K56 has been implicated in the regulation of replication, since H3K56 is transiently acetylated during the S phase to prevent hazards by DNA damage during the S phase. *rtt109* null mutants exhibit synthetic genetic interactions with mutations in Pol30 (proliferating cell nuclear antigen (PCNA)), Pol1p (DNA polymerase α), Orc2p, and Ccd45p, all of which are involved in DNA replication.

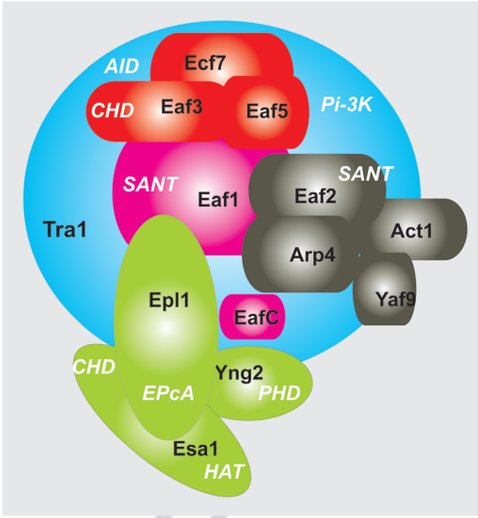


Fig. 5.2 Scheme of the NuA4–HAT complex. The recruitment module (in blue) with inserted subunits targets Esa1p-dependent acetylation to specific chromosomal loci that interact with transcription factors. The "piccolo" nucleosomal HAT module (in green) is anchored to the recruitment module; it mediates global chromatin acetylation. AID, activator-interacting domain; PI-3K, phosphatidylinositol-3-kinase; EPcA, enhancer of polycomb domain A; PHD, plant homeodomain finger; CHD, chromodomain, SANT, Swi3–Ada2–NcoR–TFIIIB domain; HAT, acetyltransferase domain.

SAGA (Spt–Ada–Gcn5 acetyltransferase; Figure 5.3) is a chromatin-modifying complex that contains two distinct enzymatic activities, Gcn5p and Ubp8p, through which it

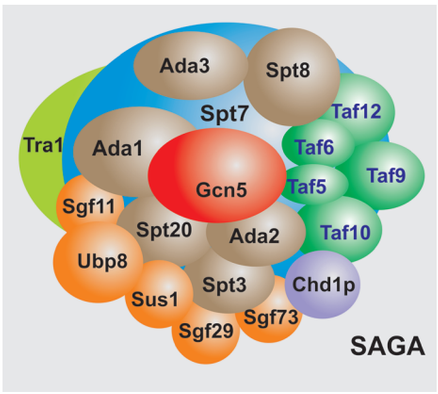


Fig. 5.3 Scheme of SAGA and its subcomplexes. DUBm components in orange.

acetylates and deubiquitinates histone residues, respectively, thereby enforcing a pattern of modifications that is decisive in regulating gene expression. The yeast SAGA complex is composed of 21 widely conserved proteins that are organized into functional submodules (Table 5.1). Some subunits of the complex have been well established by both biochemical and genetic studies, such as Ada, Spt, and a subset of TAFs, including the essential Tra1 protein, as reviewed by Daniel and Grant (2007). *SPT7* encodes a subunit of SAGA required for assembly of Spt8p into SAGA. In addition, Spt7p is required for normal amounts of two other SAGA components required for SAGA integrity, Spt20p and Ada1p, suggesting that Spt7p plays a critical role in SAGA complex formation. In addition to these subunits, new yeast SAGA modules have been discovered by biochemical approaches:

i) DUBm, the **deubiquitination module** (composed of Ubp8p, Sgf11p, Sgf29p, Sgf73p, and Sus1p) for the C-terminal ubiquitinylated K123 residue of H2B that is essential for the *trans*-tail methylation of H3 and is also required for optimal gene activation. Sus1p has significant functional links to two E2 ligases, Ubc11p and Ubc4p, as well as to the E3 ligase Ris1p/Uls1p, which implies that, in addition to being a subunit of the DUBm, it might also function as a common adaptor for both chromatin protein ubiquitination and deubiquitination.

ii) It has recently been shown that the stable association between FACT and transcribed chromatin depends on the ability to form ubiquitinated H2B, which has a role in nucleosome dynamics during transcriptional elongation (*cf.* Histone Ubiquitinylation, below). Further, there is strong evidence for a role of the SAGA complex **during transcription elongation**. Yeast SAGA has been shown not only to contact upstream activating sequences (UASs), but also to localize to the coding sequences cotranscriptionally, so that histone acetylation by Gcn5p promotes nucleosome expulsion, thereby enhancing the processivity of RNA polymerase II during elongation (Govind *et al.*, 2007). The association of SAGA with coding sequences is dependent on phosphorylation of the C-terminal domain (CTD) tail of the largest subunit of RNA polymerase II on Ser5, indicating that SAGA or some of its subunits might interact with actively transcribing RNA polymerase II during elongation.

iii) Strong evidence that gene expression and **transport of mRNA** out of the nucleus (packaged into messenger ribonucleoproteins (mRNPs) and exported through the nuclear pore complex (Iglesias and Stutz, 2008)) are tightly coupled was provided by the identification of yeast Sus1p as a subunit of both the SAGA complex and the TREX-2 complex (Pascual-Garcia and Rodriguez-Navarro, 2009).

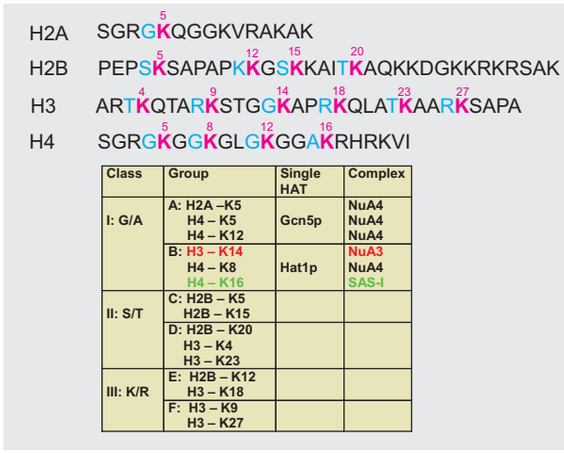
The aforementioned newly discovered components were shown to be also present in human SAGA, which confirms the conservation of the SAGA complex throughout evolution. In addition, another complex known as ATAC, which also contains GCN5, has been recently identified in humans (Guelman *et al.*, 2009; Wang *et al.*, 2008).

The six-subunit **elongator complex** is a major HAT component of the RNA polymerase II holoenzyme responsible for transcriptional elongation; there are two discrete subcomplexes (see Table 5.1). Elongator can acetylate core histones as well as nucleosomal substrates, predominant sites being H3K14 and H4K8. Although only subunit Elp5p/Iki1p is essential for growth, deletion of the other individual subunits causes significantly altered mRNA expression levels for many genes.

It seems worthwhile mentioning that recently a new nomenclature for chromatin-modifying enzymes has been introduced (Allis *et al.*, 2007).

Right signal in right place. The involvement of HATs in multiple processes seems to depend on their **substrate specificity**, for which several models have been proposed. One of these models invokes the primary sequence next to the targeted lysine residue. Figure 5.4 shows the "two-step classification" model, with the lysine residues at the N-termini of the four histones that can be acetylated. According to the nature of their left neighbors (G/A; S/T; K/R), these sites are typified into three classes, which are each subdivided into two groups on the basis of similarities in flanking amino acid sequences (Fukuda *et al.*, 2009).

Allocation of lysine specificity among members of the MYST family of HATs is different: in all, six lysines are acetylated *in vivo* by these HATs, but it appears that the "two-step classification" has to be replaced by a model that assumes that the combination of subunits in the HAT complexes dictates substrate specificity. In yeast, for example, the catalytic subunits within the MYST family complexes are Esa1p (in NuA4), Sas2p (in SAS-I), and Sas3p (in NuA3). Although



Class	Group	Single HAT	Complex
I: G/A	A: H2A –K5 H4 – K5 H4 – K12	Gcn5p	NuA4 NuA4 NuA4
	B: H3 – K14 H4 – K8 H4 – K16	Hat1p	NuA3 NuA4 SAS-I
II: S/T	C: H2B – K5 H2B – K15		
	D: H2B – K20 H3 – K4 H3 – K23		
III: K/R	E: H2B – K12 H3 – K18		
	F: H3 – K9 H3 – K27		

Fig. 5.4 Histone and site specificity of HATs and complexes. (After Fukuda *et al.*, 2009.)

individually they could acetylate six class I residues in free histones, the multisubunit complexes are required for modifications in the nucleosomal context in the following way: SAS-I for H4K16, NuA3 for H3K14, and NuA4 for H2AK5 and H4K5/8/12 (Figure 5.4).

Histone acetylation is prone to trigger further reactions involving chromatin structure, and **proteins that interact with histones** in an acetylation-dependent manner are recruited to specific acetylated residues. Structural domains that specifically recognize acetylated histones are the bromodomains of HATs that recognize specific acetyllysine residues in histones, but also in several mammalian nonhistone proteins (such as HIV Tat, p53, c-Myb, and MyoD). Additionally, bromodomains are found in many chromatin-associated proteins. The association of bromodomains and acetylated histones probably stabilizes the acetylation state and/or regulates the activities of bromodomain-containing chromatin factors. Through its bromodomain, yeast Gcn5p interacts with K16-acetylated H4 to coactivate transcription. Swi2p/Snf2p as well as Spt7p are recruited to acetylated histones for chromatin remodeling. Rsc4p is recruited to K14-acetylated H3 for chromatin remodeling, whereas Bdf1p interacts with acetylated H3 and H4 histones to establish antisilencing. Other domains, such as the chromodomain and WD40 domain, have been shown to specifically recognize modified histones.

Histone Deacetylation

How to get rid of acetyl residues? Histone acetylation is a reversible process; the removal of acetyl residues is achieved by **histone deacetylases (HDACs)**. Several HDACs have been isolated from yeast that catalyze the deacylation reaction (e.g., Peterson, 2002). HDAC families include the HDAC I class, resembling yeast Rpd3p, and the HDAC II class that is similar to yeast Hda1p. Rpd3p interacts with Sin3p and Ume1p to yield two complexes, a larger one (Rpd3L) of 1.2 MDa in size and a smaller one (Rpd3S) of 0.9 MDa in size, both of which exist also in other fungi and fulfill different functions (Table 5.2). The yeast HDACs Hos1p and Hos2p are more similar to Rpd3p (class I HDAC), while Hos3p is

Table 5.2 Yeast histone deacetylase complexes (HDACs).

Class	Complex	Component(s)	Function
I	RPD3L (large)	Rpd3p	deacetylates all four histones; transcriptional repression/activation; mediates heat stress response; involved in Sir2-mediated silencing and replication origin firing
		Sin3p	
		Ume1p	
		Dot6p	
	RPD3S (small)		interacts with Set2p-methylated histones
		Hos1p	HDAC; involved in transcriptional regulation
		Hos2p	HDAC; involved in transcriptional regulation
II	HDA complex	Hda1p	catalytic subunit of histone H2B and H3 deacetylase; transcriptional repression
		Hda2p	subunit, forms heterodimer with Hda3p
		Hda3p	subunit, forms heterodimer with Hda2p
		Hos3p	HDAC; involved in transcriptional regulation
III	SET3C		represses early/middle sporulation genes, including key meiotic regulators: Ime2 protein kinase and Ndt80 transcription factor; mammalian analog is HDAC3/SMRT complex
		Hst1p	catalytic subunit, NAD^+-dependent HDAC
		Set3p	defining member of SET3C complex
		Snt1p	subunit interacting with Sif2p
		Sif2p	WD40 repeat-containing subunit; antagonizes telomeric silencing; binds specifically to the Sir4p N-terminus
		Cpr1p	cyclophilin
		Hos2p	HDAC subunit
		Hos4p	subunit
	SIR, RENT	Sir1p	binds SIR complex to ORC complex
		Sir2p	NAD^+-dependent HDAC; cooperates with Net1p/Cdc14p in RENT complex
		Sir3p	structural subunit of SIR complex
		Sir4p	structural subunit of SIR complex
		Hst1p	NAD^+-dependent HDAC; essential subunit of the Sum1p/Rfm1p/Hst1p complex required for ORC-dependent silencing and mitotic repression
		Hst2p	cytoplasmic member of Sir2 family; modulates nucleolar silencing
		Hst3p	members of Sir2 family; involved in telomeric silencing, cell cycle progression, radiation resistance, genomic stability, and short-chain fatty acid metabolism
		Hst4p	

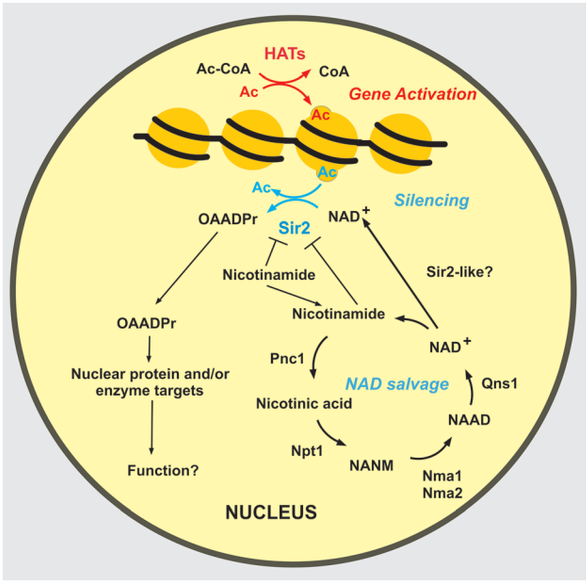


Fig. 5.5 Action of Sir2p in deacetylation.

more closely related to Hda1p (class II HDAC). Both class I and class II deacetylase use coenzyme A as a cofactor. A third group (class III HDACs) includes the SET3C complex (with Set3p as the defining member and Hst1p as a catalytic subunit (Pijnappel *et al.*, 2001)) as well as the "Sirtuin" family (Blander and Guarente, 2004; Sauve *et al.*, 2006), whose founding member is the Sir2p protein from *S. cerevisiae*. Both complexes are structurally unrelated to the other two HDAC families (which are zinc-dependent) and have the unusual property of requiring NAD^+ as a cofactor in the deacetylation reaction (Peterson, 2002; Robyr *et al.*, 2002; Denu, 2003).

Unique HDACs. The NAD^+-dependent deacetylases work on histones and/or other proteins. They catalyze a unique reaction in which the cleavage of NAD^+ and deacetylation is coupled to the formation of *O*-acetyl-ADP-ribose (OAADPr) – a novel metabolite (Figure 5.5). In this reaction, nicotinamide is liberated from NAD^+ and the acetyl group of the substrate is transferred to cleaved NAD^+, generating OAADPr.

Deacetylation may either be promoter-targeted to inactivate certain genes or exert a global effect (i.e., to restore the epigenetic pattern after replication). The involvement of the **SIR complex** in generating boundaries in chromatin structure and the participation of Sir2p in the **RENT** (regulator of nucleolar silencing and telophase exit) complex are explained in Section 5.1.3.3. Examples of transcriptional regulation by SIR are given in Section 9.3.

Histone Methylation

Appending methyl groups. Histone methylation (Trievel, 2004; Fuchs, Laribee, and Strahl, 2009) is chemically more stable and for a long time methylation was believed to be irreversible, since no histone demethylases had been isolated; this situation has changed recently. The modification adds methyl groups either to a lysine $\varepsilon\text{-}NH_3^+$ group, which (according to its structure) can be mono-, di-, or trimethylated, or to an arginine residue, which can accommodate two methyl groups. Histone methylation is carried out by **histone methyltransferases (HMTs)** that use *S*-adenosylmethionine (SAM) as coenzyme; six SET homologs of the mammalian factors are found in yeast (Set1p–Set6p). All of them contain a so-called SET domain; however, only three of the members have been functionally identified. Set1p is the catalytic subunit of the six-unit SET1C (**COMPASS**) complex (Table 5.3). Lysine methylation of histones in budding yeast has only been identified at a few locations, but each one is handled by a specific methyltransferase. Histone H3 Lys4 (H3K4) is modified by Set1p (Briggs *et al.*, 2001) and histone H3 Lys36 (H3K36) by Set2p (Strahl *et al.*, 2002), whereas H3 Lys79 (H3K79) is methylated in its core region by Dot1p (Feng *et al.*, 2002). The discovery of the Dot1 family of HMTs came as a surprise and only when overexpression of Dot1p led to the disruption of telomeric silencing.

Unlike acetylation, the charge of the methylated histones will not change. However, similar to acetylation, methylated residues will recruit additional factors binding to them through a number of protein domains. These domains include chromodomains (Jacobs and Khorasanizadeh, 2002) and PHD domains (Li *et al.*, 2006), which bind methyllysine, and Tudor domains, which recognize both methyllysine and methylarginine. As a consequence, these factors lead to remodeling of chromatin structure, thus inducing complex patterns of gene expression. For example, methylation at H3K4, H3K36, and H3K79, is found associated with active transcription (*cf.* Section 9.3).

Arginine methylation in yeast has been identified on histone H4 (H4R3) and on histone H3 (H3R2) (Kirmizis *et al.*, 2007); the prevailing enzyme is the arginine methyltransferase Hmt1p. H4R3 possibly has a role in gene silencing.

Histone demethylation in yeast has been a mystery for a long time, but recently Jhd2p, a JmjC domain family histone demethylase specific for H3K4, has been found (Liang *et al.*, 2007). The enzyme removes two or three methyl groups that were specifically added by the Set1p methyltransferase. The protein levels are regulated by Not4p, an E3 ubiquitin ligase, which mediates the polyubiquitination and degradation of Jhd2p. The H3K36 methylation is removed by two different demethylases, Jdh1p and Rph1p (Fang *et al.*, 2007; Kim and Buratowski, 2007).

Histone Ubiquitinylation

Anticipating ubiquitin. Ubiquitinylation (*cf.* Section 6.3) is common for modifying lysine residues, arguing that in histones it might compete with acetylation or methylation; however, this absolutely contradicts all findings. In yeast, only a single residue, Lys123 in H2B (at the C-terminal tail), has thus far been identified as a substrate: first, monoubiquitination was observed (Robzyk, Recht, and Osley, 2000), and,

Table 5.3 Yeast histone methylases (HMTs).

Complex	Components	Function
SET1C (COMPASS) HMT complex; methylates H3K4; required in transcriptional silencing near telomeres and at silent mating-type loci; contains a SET domain	Set1p	catalytic subunit
	Bre2p	subunit of SET1C; interacts with Sdc1p
	Sdc1p	subunit of SET1C; mediates interaction with Bre2p
	Shg1p	subunit of SET1C
	Spp1p	subunit of SET1C; interacts with Orc2p; PHD finger domain protein
	Swd1p	WD40 β-propeller superfamily member
	Swd2p	subunit of CPF (cleavage and polyadenylation factor) – a complex involved in RNA polymerase II transcription termination
	Swd3p	WD40 β-propeller superfamily member
Set2p	Set2p	HMT, has a role in transcriptional elongation, methylates Lys36 of H3; associates with C-terminal domain of Rpo21p; histone methylation activity is regulated by phosphorylation status of Rpo21p
SET3C	Set3p	member of the SET3C HDAC complex; repressor of early/middle meiosis-specific genes (like *IME2* and *NDT80*); contains both SET and PHD domains
	Snt1p	subunit of SET3C, interacting with Sif2p
	Sif2p	subunit of SET3C, interacting with Snt1p
	Cpr1p	cyclophilin
	Hos2p	subunit of SET3C, HDAC
	Hos4p	subunit
	Hst1p	subunit of SET3C, putative substrate of Cdc28p; NAD-dependent HDAC; nonessential subunit of SET3C
	Set4p	protein of unknown function, contains a SET domain
	Set5p	zinc finger protein of unknown function, contains one canonical and two unusual fingers in unusual arrangements; deletion enhances replication of positive-strand RNA virus
	Set6p	SET domain protein of unknown function; heterozygote deletion is sensitive to compounds that target ergosterol biosynthesis, may be involved in compound availability
DOT	Dot1p	nucleosomal histone H3K79 methylase; methylation is required for telomeric silencing, meiotic checkpoint control, and DNA damage response
	Dot4p (Ubp10)	ubiquitin-specific protease that deubiquitinates ubiquitin–protein moieties; may regulate silencing by acting on Sir4p; primarily located in the nucleus
	Dot6p	protein involved in rRNA and ribosome biogenesis; binds polymerase A and C motif; subunit of the RPD3L HDAC complex; similar to Tod6p; has chromatin-specific SANT domain; involved in telomeric gene silencing and filamentation

recently, polyubiquitination has been reported. The modification is catalyzed by the Rad6p (Ubc2p)–Bre1p ubiquitin–ligase complex, and functions in transcriptional initiation and elongation. Ubiquitination of H2BK123 is required for H3K4 methylation and H3K79 in yeast (as well as in higher eukaryotes), whereby H3K4 methylation cooperates with COMPASS and the chromatin-remodeling FACT complex in transcriptional elongation. Activation of this process strictly depends on the deubiquitination of K123 by the protease Ubp8p – a component of SAGA. We will come back with further details in Section 9.3.

In higher eukaryotes, H2A is also prone to ubiquitination, the moiety attached to a position corresponding to Lys119. This modification attracts association with a number of repressive complexes, which have been discussed elsewhere (Weake and Workman, 2008).

Histone SUMOylation Modifications by the attachment of SUMO (Smt3p in yeast) appears to be possible at several locations in all histones. In H4, SUMOylation has been detected at five (unspecified) positions; in H2B modifications are thought to occur at Lys6/7 and Lys16/17, while

on H2A SUMO is attached to Lys126. Unfortunately, little is known about the functions of SUMOylation of histones. Generally, it is accepted that SUMOylation sets repressive marks for transcription, in a way antagonizing histone acetylation.

Histone Phosphorylation

Negative charges. Phosphorylation of histones is rather scarce, but a few of these seem to be of functional significance. For example, H3S10 (phosphorylation of Ser10) supports transcription by interfering with the acetylation at H3K14; H2BS10 (effected by Ste20p) is a signal for apoptosis after oxidative stress; and phosphorylation of the C-terminal tail on H2A or of H4S1 both are linked to induction of DNA damage repair, whereby the phosphorylation at H4S1 recruits both the NuA4 HAT complex and the SWI/SNF remodeling complex to genes involved in histone acetylation or polymerase II activity (Utley *et al.*, 2005).

Histone Code

Foundation of epigenetics? Histone modifications may affect chromatin structure directly by altering DNA–histone interactions within and between nucleosomes, thus changing higher-order chromatin configuration. This approach has been termed "the direct interaction" model. An alternative model suggested that combinations of histone modifications provide an interaction surface for other proteins, which translate this so-called **histone code** (Jenuwein and Allis, 2001) into a gene expression pattern. In other words, the histone code hypothesis (and variations thereof) postulates that particular patterns of histone modification function as a signal to other proteins containing histone-binding domains, which then bind to histones in a modification-specific mode, thus recruiting chromatin-remodeling factors (Nightingale, O'Neill, and Turner, 2006). For example, when histone H2AS129 is phosphorylated by Mec1p during double-stranded DNA cleavage (Downs, Lowndes, and Jackson, 2000), NuA4 is recruited to its specific target lysine via recognition of phosphorylated H2AS129P by the NuA4 subunit Arp4 (Downs *et al.*, 2004). NuA4 subsequently acetylates histone H4. The histone code model could also easily explain how the same chemical modification can have different functional consequences depending on the respective target site (e.g., methylation of a particular residue in H3 is correlated with gene activation, while methylation of another residue in H3 results in repression and heterochromatin formation).

By contrast, histone acetylation is generally correlated with gene activation. The mechanisms for sequential modification and regulation of chromatin function are very similar to those that have been proposed for cellular signal transduction (Schreiber and Bernstein, 2002). We will elaborate on the epigenetic consequences for gene expression in more detail in Section 9.3.

5.1.1.2.2 Remodeling Chromatin Structure

Overview

Chromatin reorganization facilitates life. Remodeling complexes either change the location of a nucleosome along a particular DNA sequence (originally termed "nucleosome sliding") or create a remodeled state of the nucleosome that is characterized by altered histone–DNA interactions. The first such activity found was the Swi2/Snf2 ATPase, required both for ATP hydrolysis and for coupling the energy from ATP hydrolysis to the ATP-dependent remodeling complex SWI/SNF, whose main activity was noted to alter histone–DNA contacts within nucleosomes (Cairns *et al.*, 1994). (Originally, the SWI genes were discovered in 1984 (Stern, Jensen, and Herskowitz, 1984) and later, together with SNF genes, recognized as "global transactivators" (Winston and Carlson, 1992; Carlson and Laurent, 1994)). In yeast, mutations in the SWI2/SNF2 gene are responsible for growth defects, such as the inability to undergo mating-type switching (swi^-) and sucrose nonfermenting (snf^-), inducing faults in the expression of characteristic sets of genes.

Finally, the **SWI/SNF** multiprotein complex (composed of Swi2p/Snf2p, Swi1p, Swi3p, Snf5p, Snf11p, Snf12p, Swp82p, and Arp7p; Table 5.4) became a paradigm for chromatin remodeling (Peterson and Tamkun, 1995; Wilson *et al.*, 1996). The SWI/SNF complex is highly related to the **RSC** (remodels the structure of chromatin) complex in yeast (Cairns *et al.*, 1996), which harbors the ATPase Snf2p homolog Sth1p (Cao *et al.*, 1997; Cairns *et al.*, 1999). Interestingly, both complexes contain actin-related proteins (see below).

Meanwhile, a variety of chromatin-remodeling complexes became known, the majority of which contain an ATPase as a **central motor subunit** (except FACT), and perform critical functions in the maintenance, transmission, and expression of eukaryotic genomes by regulating the structure of chromatin (Table 5.4). Most of the remodeling complexes exert their regulatory activities in several areas. For example, the SWI/SNF chromatin-remodeling complex is involved in DNA replication, stress response, and transcription. In the latter case, the complex binds DNA nonspecifically, altering the nucleosome structure such that binding of transcription factors is facilitated. For some genes, transcriptional activators are able to direct the SWI/SNF complex to the UAS in the promoter.

Chromatin-Remodeling Complex SWI/SNF

"Switching" around. The **SWI/SNF** chromatin-remodeling complex family is evolutionary highly conserved and comprises two subfamilies of chromatin-remodeling factors: one subfamily includes yeast SWI/SNF, fly BAP, and mammalian BAF; the other subfamily comprises yeast RSC, fly PBAP, and mammalian PBAF. In mammalian cells, subunits of the SWI/SNF complex appear to act as tumor suppressors and targets for oncoproteins that disrupt the normal cell cycle control by E1A.

It may be noted that Snf2p – the founding member of the Snf2p subfamily of proteins (which in yeast includes Chd1p,

Table 5.4 Chromatin-remodeling complexes.

Complex	Components	Features
SWI/SNF complex	Swi2p/Snf2p	ATPase; central motor subunit
	Swi1p	binds promoter activation domains; can form the prion [SWI$^+$]
	Swi3p	two copies present; provides structural integrity
	Snf5p	Important for the assembly of the SWI/SNF complex and its nucleosome remodeling activities; may be involved in negative regulation of chromatin silencing
	Snf11p	two copies of Snf11p per SWI/SNF complex; interacts with an evolutionarily conserved 40-residue sequence in Snf2p
	Snf12p	homolog of Rsc6p; required for maintaining the full structural integrity of the complex; binds to sequences in the activation domain of transcriptional activator Gcn4p, thus contributing to the ability of Gcn4p to recruit the entire SWI/SNF complex to promoters
	Swp82p	two copies in the complex; abundantly expressed in many growth conditions; able to activate transcription *in vitro*
	Arp7p (Swp61p, Rsc11p)	component of both the SWI/SNF and RSC chromatin-remodeling complexes; actin-related protein
	Arp9p	C-termini of Arp7p and Arp9p are both required for association of the Arp7p/Arp9p heterodimer with the RSC complex
	Taf14p	component of a number of different complexes; contains YEATS domain
	Rtt102p	component of both the SWI/SNF and RSC chromatin-remodeling complexes, suggested role in chromosome maintenance; possibly weak regulator of Ty1 transposition
RSC complex	Sth1p	essential ATPase component; required for expression of early meiotic genes; essential helicase-related protein homologous to Snf2p
	Sfh1p	required for cell cycle progression and maintenance of proper ploidy; phosphorylated in the G_1 phase of the cell cycle; Snf5p paralog
	Spt6p	transcription elongation factor
	Rsc1p	contains two essential bromodomains, bromoadjacent domain (BAH) and AT hook; required for expression of mid/late sporulation-specific genes
	Rsc2p	required for expression of mid/late sporulation-specific genes; involved in telomere maintenance
	Rsc3p	absolutely required for maintenance of proper ploidy, regulation of ribosomal protein genes and the cell wall/stress response; highly similar to Rsc30p
	Rsc4p	found close to nucleosomal DNA; displaced from the surface of nucleosomal DNA after chromatin remodeling
	Rsc6p	essential for mitotic growth; homolog of SWI/SNF subunit Snf12p
	Rsc7p (Npl6p)	interacts with Rsc3p, Rsc30p, Ldb7p, and Htl1p to form a module important for a broad range of RSC functions; involved in nuclear protein import and maintenance of proper telomere length
	Rsc8p (Swh3p)	essential for viability and mitotic growth; homolog of Swi3p, but does not activate transcription of reporters
	Rsc9p	DNA-binding protein involved in synthesis of rRNA and in transcriptional repression/activation of genes regulated by TOR (target of rapamycin) pathway
	Rsc11p (Arp7p)	actin-related protein; also in SWI/SNF
	Rsc12p (Arp9)	actin-related protein
	Rsc14p (Ldb7p)	interacts with Rsc3p, Rsc30p, Npl6p, and Htl1p to form a module important for a broad range of RSC functions
	Rsc30p	required for regulation of ribosomal protein genes and the cell wall/stress response; highly similar to Rsc3p
	Rsc58p (Htl1p)	functions in transcriptional regulation and elongation, chromosome stability, and establishing sister chromatid cohesion; involved in telomere maintenance
	Rtt102p	also subunit of SWI/SNF
SWR1 complex	Swr1p	Swi2/Snf2-related ATPase, structural component of SWR1 complex, which exchanges histone variant H2A.Z (Htz1p) for chromatin-bound histone H2A
	Swc4p (Eaf2p)	incorporates Htz1p into chromatin; component of the NuA4 HAT complex
	Swc3p	unknown function; required for formation of nuclear-associated array of smooth endoplasmic reticulum
	Swc5p	unknown function
	Swc7p	unknown function
	Yaf9p	subunit of both the NuA4 histone H4 acetyltransferase complex and the SWR1 complex, may function to antagonize silencing near telomeres; interacts directly with Swc4p; has homology to human leukemogenic protein AF9; contains a YEATS domain
	Vps71p	component of the SWR1 complex
	Vps72p	Htz1p-binding component of the SWR1 complex; required for vacuolar protein sorting

Table 5.4 (*Continued*)

Complex	Components	Features
	Arp6p	actin-related protein that binds nucleosomes
	Htz1p	histone variant H2A.Z
ISW1		imitation switch (ISWI) class of ATP-dependent chromatin-remodeling complexes
ISW1a	Isw1p	ATPase, forms complex with Ioc3p
	Ioc3p	represses transcription initiation by specific positioning of a promoter proximal dinucleosome
ISW1b	Isw1p	ATPase, forms complex with Ioc2p and Ioc4p that regulates transcription elongation
	Ioc2p	contains PHD finger motif
	Ioc4p	contains PWWP motif
ISW2 (yCHRAC)		ATP-dependent DNA translocase involved in chromatin remodeling
	Isw2p	ATPase component, forms a complex with Itc1p
	Itc1p	required for repression of **a**-specific genes, *INO1*, and early meiotic genes during mitotic growth
	Dls1p	involved in inheritance of telomeric silencing
	Dbp4p	involved in both chromosomal DNA replication and inheritance of telomeric silencing; also subunit of DNA polymerase ε
INO80 complex	Ino80p	ATPase, subunit of a complex containing actin and several actin-related proteins; has chromatin-remodeling activity (and 3′ → 5′ DNA helicase activity *in vitro)*; has a role in modulating stress gene transcription
	Taf14p	component of several chromatin-remodeling complexes
	Act1p	actin
	Arp4p	nuclear actin-related protein in remodeling complexes
	Arp5p	nuclear actin-related protein in remodeling complexes
	Arp8p	nuclear actin-related protein in remodeling complexes
	Ies1p	subunit
	Ies3p	subunit
	Ies4p	target of the Mec1p/Tel1p DNA damage signaling pathway
	Nhp10p	protein related to mammalian HMG proteins
	Rvb1p	essential protein involved in transcription regulation (pontin)
	Rvb2p	essential protein involved in transcription regulation (reptin)
FACT		abundant nuclear complex; required for transcription elongation on chromatin templates; destabilizes the interaction between the H2A/H2B dimer and the H3/H4 tetramer, thus reorganizing nucleosome structure; may play a role in DNA replication and other processes that traverse chromatin
	Pob3p	binds to nucleosomes via Nhp6
	Spt16p	required for the maintenance of chromatin structure during transcription, avoiding transcription of cryptic promoters
	Nhp6Ap	HMG proteins that bind to and remodels nucleosomes; involved in recruiting FACT and other chromatin remodeling complexes to chromosomes; homologous to mammalian HMGB1 and HMGB2
	Nhp6Bp	

Isw1p, Isw2p, Ino80p, Mot1p, Rad5p, Rad16p, Rad54p, Rdh54p, Spt20p, Sth1p, Swr1p, and Uls1p) – is part of the nucleic acid-dependent ATPase and **helicase superfamily** (*cf.* Section 6.5). Although several family members have been shown to exhibit DNA-stimulated ATPase activity, no DNA helicase activity has been ascribed to any member of the Snf2 subfamily. Snf2p is also similar Spt7p. Snf2p is functionally interchangeable with homologs from other species and chimeras were also found to be active.

Swi3p as a subunit is present in two copies per complex and required for maintenance of the full structural integrity of the SWI/SNF complex. Swi3p is involved in transcription of a diverse set of genes, including HO and Ty retrotransposons. It is also required for normal mating-type switching and recruitment of SWI/SNF to promoters by Gcn4p. Swi3p has two domains that are essential for its function – a SWIRM domain (predicted to mediate specific protein–protein interactions) and a SANT domain. *swi3* null mutants are viable, but grow slowly on glucose, are inositol auxotrophs, and are unable to grow aerobically on maltose, galactose, or raffinose. Further, *swi3* mutants are defective in mating-type switching and sporulation.

Snf5p is important for the assembly of the SWI/SNF complex and its nucleosome-remodeling activities; it may be involved in the negative regulation of chromatin silencing. Snf5p is required for the normal expression of all histone genes, including HTA1 and HTB1. Hir1p and Hir2p bind Snf5p and target it, together with the SWI/SNF complex, to the HTA1–HTB1 locus. Snf5p interacts with Taf14p. *snf5* null mutants are viable, but display reduced growth on glucose and sucrose, are unable to grow on raffinose, galactose, or glycerol, and are hypersensitive to lithium and calcium ions. Snf5p is similar to Sfh1p, *Drosophila* SNR1, *Schizosaccharomyces pombe* Snf5p, and *Arabidopsis thaliana* BSH,

which can partially complement the defects seen in *snf5* null mutants. The human homolog of Snf5p (SMARCB1) is a tumor suppressor, mutation of which is associated with oncogenesis.

A component that the SWI/SNF complex shares with a number of other complexes (such as transcription factor TFIID, Mediator, the nucleosomal histone H3 acetyltransferase (NuA3), and INO80) is Taf14p. **Taf14p** is also a component of the transcription factor TFIIF complex, but is less tightly associated with TFIIF than its other components (Tfg1p or Tfg2p) and is not essential for TFIIF function. Actually, SWI/SNF contains three copies of Taf14p, which directly interact with catalytic proteins Tfg1p (TFIIF) and Sth1p (in the RSC complex), and appears to interact with catalytic subunits (Taf2p, Ino80p, and Sas3p) of other complexes that participate in RNA polymerase II-mediated transcription initiation (like TFIID, INO80, and NuA3 complexes). Taf14p is responsible for efficient transcription in yeast, suggesting that it takes a common regulatory function in each of these complexes.

Other important features of Taf14p are involvement in bud morphogenesis, formation of mating projection, actin organization, localization of Spa2p (the scaffold for cell wall integrity signaling components), and negative regulation of chromatin silencing. Taf14p may also affect the cell cycle arrest functions of Rad53p and Mec1p. Characteristic for Taf14p is a so-called YEATS domain, which is also found in Yaf9p and Sas5p.

Essential components of both the SWI/SNF and RSC chromatin-remodeling complexes are the are nuclear **actin-related proteins** Arp7p and Arp9p, which form a stable heterodimer. For formation of the RSC complex, the C-termini of Arp7p and Arp9p are both required for association of the heterodimer with the complex. Depending on the genetic background, *arp7* and *arp9* null mutants are each either nonviable or show greatly impaired growth with mutant phenotypes similar to those seen in *snf2* null mutants, such as an inability to grow on nonfermentable carbon sources. Genetic analyses have indicated that the Arp7p/Arp9p heterodimers may also cooperate with Nhp6ap and Nhp6bp (see below) to facilitate proper chromatin architecture.

Chromatin-Remodeling Complex RSC As indicated above, the RSC complex is related to the Swi2/Snf2 complex. The ATPase subunit has been termed Sth1p (Snf two homolog) and exhibits helicase activity (Laurent *et al.*, 1992; Du *et al.*, 1998). Similar to the Swi2/Snf2 complex, the RSC complex is involved in chromatin remodeling. It is particularly required for the expression of early meiotic genes in yeast (Yukawa *et al.*, 1999) and also for kinetochore function during chromosome segregation (Hsu *et al.*, 2003). One component, Sfh1p (Snf five homolog) is required for cell cycle progression and maintenance of proper ploidy. Further, there are 13 RSC-specific subunits (Table 5.4), one of which (Rsc1p) contains two bromodomains. As in the SWI/SNF complex, Rtt109p is also a subunit of the RSC complex.

Chromatin-Remodeling Complex SWR1 SWR1 is a chromatin-remodeling complex, which is active in exchanging chromatin-bound histone H2A against the variant H2A.Z histone (Htz1p) (Krogan *et al.*, 2003; Mizuguchi *et al.*, 2004). The structural component is an ATPase related to the Swi2/Snf2 protein; its name is derived from a mammalian paralog, "sick with Rat8 ts." The complex carries the histone Htz1p with it bound to and transferred by Swc4p (Wu *et al.*, 2005).

Chromatin Remodeling by ISWI The family of "imitation switch" (ISWI)-type ATPases was identified on the basis of their similarity to Swi2/Snf2 and belong to the "sliding-type" of remodeling activities (Stern, Jensen, and Herskowitz, 1984; Winston and Carlson, 1992; Carlson and Laurent, 1994; Peterson and Tamkun, 1995; Wilson *et al.*, 1996; Morillon *et al.*, 2003a). It may be noted that three complexes – CHRAC (chromatin accessibility complex), ACF (ATP-utilizing chromatin assembly and remodeling factor), and NURF (nucleosome remodeling factor) – were biochemically isolated from *Drosophila melanogaster* and contain ISWI as the ATPase component. Nucleosome-remodeling ATPases of the CHD type are characterized by the presence of a pair of the so-called chromodomains. In vertebrates, several members of the family have been identified. Two ISWI homologs, Isw1p and Isw2p, are present in yeast. There are several possibilities for these ATPases to complex with different partners. (i) Isw1p is able to form complexes (Isw1b) with Ioc2p (characterized by a PWWP motif) and/or Ioc4p (characterized by a PHD finger domain) that coordinate transcription elongation and termination. (ii) Another complex (Isw1a) containing Iswp1 with Ioc3p is required for promoter inactivation by preventing polymerase II from associating with the promoter (repression of transcription initiation).

The **ISW2** (or yCHRAC) complex, formed by Isw2p with Itc1p, Dsl1p, and Dpb4p, is required for repressing **a**-specific genes, *INO1*, and early meiotic genes during mitotic growth to repress these during vegetative growth (Ehrenhofer-Murray, 2004; Morillon *et al.*, 2003; Mellor and Morillon, 2004). Dbp4p is a subunit of ISW2 that is involved in chromosomal DNA replication as well as in the inheritance of telomere silencing; it is also a subunit shared by DNA Pol2 ε (*cf.* Table 5.5).

Chromatin-Remodeling Complex INO80 Yet another remodeling activity, INO80, is modulated by inositol phosphates (Ebbert, Birkmann, and Schüller, 1999; Jones and Divecha, 2004). The ATPase subunit of this complex, containing actin and several actin-related proteins, is Ino80p (*cf.* Table 5.4). The complex has chromatin-remodeling activity (and $3' \rightarrow 5'$ DNA helicase activity *in vitro*). In particular, INO80 has a role in modulating stress gene transcription and is involved in DNA damage repair (Klopf *et al.*, 2009); it will also act as a nucleosome spacing factor (Udugama, Sabri, and Bartholomew, 2011). The significance of phosphatidylinositol phosphates (PIPs) has been outlined in Section 3.4.

Chromatin Reorganization by FACT

FACTs count. The **FACT** (facilitates chromatin transactions) complex is a heterodimer consisting of the highly conserved subunits Pob3p and Spt16p (review: Formosa, 2008). Spt16p has been characterized as a transcription elongation factor required for the maintenance of chromatin structure during transcription, thus avoiding transcription to (cryptic) promoters within the genes (Bortvin and Winston, 1996; Hartzog *et al.*, 1998; Kaplan, Laprade, and Winston, 2003). In higher eukaryotes, the Pob3p (Pol1-binding) homolog comprises a high-mobility group (HMG) DNA-binding motif that effects chromatin binding, whereas in *S. cerevisiae*, chromatin association of the complex is mediated by the small HMG-box protein Nhp6p (encoded by two nearly identical genes, *NHP6A* and *NHP6B*) (Formosa *et al.*, 2001), whereby Nhp6p binds to histones prior to Pob3p/Spt16p binding. Although Pob3p, Spt16p, and Nhp6Ap/Nhp6Bp do not form a stable heterotrimer, the Nhp6 protein is necessary for activity of the FACT complex. Nhp6p contacts nucleosomal DNA without sequence specificity and bends it sharply. There is approximately one molecule of Nhp6Ap present for every one to two nucleosomes and 1/10th as much of Nhp6Bp, consistent with the observed 3- to 10-fold difference in mRNA levels. Many experimental results indicate that Nhp6p also interacts with other known chromatin-remodeling activities (SWI/SNF, RSC, Ssn6p, and Spt6p) by loosening or remodeling the structure of the core nucleosome. Nhp6p likely serves to guide the complexes to appropriate places within the chromatin. Nhp6p activity has been shown to contribute to DNA replication (Vandemark *et al.*, 2006; Han *et al.*, 2010), by the formation and correct placement of preinitiation complexes (preinitiation complex PICs) for certain genes transcribed by either RNA polymerase II or III, including the essential U6 small nuclear RNA (snRNA). Nhp6p is also implicated in DNA mismatch repair (MMR): the MutS-α complex, composed of Msh2p and Msh6p, colocalizes with Nhp6p and DNA containing mismatches both as part of the FACT complex and independently of FACT as well.

The FACT complex – in a mechanism distinct from ATP-dependent chromatin remodeling – is able to alter chromatin structure without the requirement for ATP hydrolysis (Xin *et al.*, 2009). Recent work suggests that the alterations in chromatin structure induce nucleosome reorganization. However, FACT does not seem to have a role in chromatin reassembly within promoters after transcription initiation. The role of FACT in transcriptional elongation is discussed in more detail in Section 9.3.

5.1.2 Centromeres

Chromosomes need one center each. The **centromeric DNA sequences** in all yeast chromosomes share a common substructure (Fitzgerald-Hayes, Clarke, and Carbon, 1982; Hieter *et al.*, 1985), which extends over only some 200 bp, contrary to the much larger centromeres in *S. pombe* or mammalian cells, where they occupy some 200 kb. The centromere sequences from *S. cerevisiae* can be subdivided into three distinct regions, which differ in base composition (Panzeri *et al.*, 1985). The central part containing the consensus sequence AAWTWART*CACRTG*ATAWAWWT (centromere DNA element I (CDEI)) represents the binding site (CACRTG) for a basic helix–loop–helix (bHLH) DNA-binding protein, the centromere-binding factor (Cbf1p), which was also shown to play a possible role as a transcription factor (Bram and Kornberg, 1987). Cbf1p is required for nucleosome positioning at the motif; it targets the remodeling complex Isw1p to the DNA. Centromeric DNA sequences are the sites of kinetochore formation (Lechner and Ortiz, 1996) and chromosome attachment to mitotic and meiotic spindles. The kinetochore – one each for each sister chromatid after chromosome duplication – is composed of protein assemblies that can be broadly classified into inner, central, or outer kinetochore complexes; more than 40 different factors have been characterized that build up these structures. The kinetochores are the distal points of "nuclear" **microtubule attachment** during chromosome segregation, while the proximal attachment sites for the microtubules are localized to inner plaque of the spindle pole body (SPB) (*cf.* also Chapter 7).

5.1.3 Replication Origins and Replication

5.1.3.1 Initiation of Replication

Make two out of one. A multitude of insights into the basal cellular processes of chromosome replication were gained from the studies employing yeast as a model system. As early as in 1979, the laboratory of Ron Davis isolated and characterized a yeast chromosomal replicator (Stinchcomb, Struhl, and Davis, 1979) that turned out to be a comparably short segment of DNA. Such sequences functioning as **autonomous replication origins** (autonomous replication sequences (ARSs)) – also suitable for autonomously replicating yeast plasmid vectors (Struhl *et al.*, 1979; Stinchcomb *et al.*, 1980) – were found not only to be present within the centromeric regions, but also to occur in similar copies along all yeast chromosomes at about 30 kb intervals (Chan and Tye, 1980; Newlon, 1988; Newlon and Theis, 1993). Chromosomal ARS and centromere (CEN) elements were observed to bind specifically to the yeast nuclear scaffold (Amati and Gasser, 1988).

In contrast to the complex and highly conserved replicators present in prokaryotes and viruses, under study in Bruce Stillman's laboratory (Tamanoi and Stillman, 1983; Stillman and Gluzman, 1985), no conserved sequences have been detected in the sequences that autonomously replicate in yeast with the exception of a single 11-bp element called the ARS consensus sequence (ACS) (Broach *et al.*, 1983). This element was found to be essential, but not sufficient, for replicator function (Deshpande and Newlon, 1992; Rivier and Rine, 1992; Huang and Kowalski, 1993). A subset of ARSs colocalizes with origins of replication in the chromosome

(Fangman and Brewer, 1991), suggesting a functional link between chromosomal replicators and the autonomous replication property of these sequences. Detailed analyses of ARS sequences by the Stillman group led to the identification of four short sequences (A, B1, B2, and B3) that collectively are sufficient for efficient replicator function (Marahrens and Stillman, 1992; Rao, Marahrens, and Stillman, 1994): element A contains the above conserved ARS consensus sequence and is bound by the **origin recognition complex (ORC)** – the initiator protein of replication (Bell, Kobayashi, and Stillman, 1993); element B3 was identified as a protein binding site for **ARS-binding factor 1** (Abf1p) that also functions as a transcription factor at a large variety of promoters (Diffley and Stillman, 1988, 1989). Later on, it was shown that Abf1p is a component of the nucleotide excision repair (NER) complex (Reed *et al.*, 1999; Ellison and Stillman, 2003). Abf1p levels are abundant in the cell, so that Abf1p-binding sites in the genome are occupied *in vivo* under all conditions studied thus far. On the other hand, it has been shown that Abf1p is capable of repressing its own transcription by binding to a consensus site in the *ABF1* promoter. The binding activity of Abf1p is stimulated by Cdc6p. Abf1p can be phosphorylated at multiple sites, partially through the action of serine/threonine kinase, whereby the extent of phosphorylation depends on growth conditions and carbon source. Changes in Abf1p phosphorylation have been shown to correlate with regulation of expression of the Abf1p target gene *COX6*, linking Abf1p phosphorylation with carbon-source control of *COX6*. Dephosphorylation requires the presence of functional Cyc8p. Nuclear import of Abf1p is dependent on the Ran-GEF (guanine nucleotide exchange factor) Srm1p, but Abf1p can also be imported via importin Pse1p, suggesting that import is mediated by more than one pathway. Export of Abf1p mRNA is dependent on export factor Yra1p.

Along with the analysis of ARS functions, the Stillman group and many other renowned researchers investigated the **replication machinery** and the mechanism of replication in yeast. This was done in parallel in humans, and resulted in the characterization of the DNA polymerases and other components involved in DNA replication (Prelich *et al.*, 1987; Heintz and Stillman, 1989; Smith and Stillman, 1989; Tsurimoto and Stillman, 1989; Din *et al.*, 1990; Brill and Stillman, 1991; Smith and Stillman, 1991; Estes, Robinson, and Eisenberg, 1992; Fien and Stillman, 1992; Cullmann *et al.*, 1995); details of the process are still under study.

The first event in DNA replication is the binding of the ORC to multiple ARS sequences. These "prereplication" (or preinitiation) complexes (**pre-RCs** or PRCs) are assembled during the M and G_1 phase, whereby this binding persists throughout the cell cycle. The six subunits of the ORC complex, which was also shown to be involved in transcriptional silencing, were isolated and characterized (Diffley and Cocker, 1992; Foss *et al.*, 1993; Micklem *et al.*, 1993; Bell *et al.*, 1995; Loo *et al.*, 1995; Rao and Stillman, 1995; Li *et al.*, 1998; Du and Stillman, 2002). ORC is an ATP-dependent DNA-binding protein complex, the subunits of which are encoded in yeast by *ORC1–ORC6*. ORC directs DNA replication throughout the genome and is required for its initiation. Both Orc1p and Orc5p bind ATP, but only Orc1p exhibits ATPase activity. The binding of ATP by Orc1p is required for ORC binding to DNA and is essential for cell viability. The stability of ORC as a whole depends on ATP binding by Orc5p as well as on Orc6p, once the pre-RC has been formed. ORC homologs have been characterized in various other eukaryotes, including fission yeast, insects, amphibians, and humans.

Most importantly, in the late M phase the ATP-dependent protein **Cdc6p** is recruited by the ORC, together with Tah11p/Cdt1p, which in turn promotes loading of the **minichromosome maintenance (MCM)** complex on to chromatin (Liang, Weinreich, and Stillman, 1995; Cocker *et al.*, 1996; Williams, Shohet, and Stillman, 1997; Zou, Mitchell, and Stillman, 1997; Raghuraman *et al.*, 2001; Stillman, 2001; Weinreich *et al.*, 2001; Stillman 2005); therefore, Cdc6p is called the "**loading factor**." The Mcm2p–Mcm7p family is a group of six proteins that are highly conserved in all eukaryotes, with homologs having also been identified in Archaea. Three of the genes (*MCM4/CDC54*, *MCM5/CDC46*, and *MCM7/CDC47*) were originally identified as CDC mutants. In *S. cerevisiae*, each of the six Mcm2–7 proteins is essential for viability. Both *CDC6* mRNA and protein levels peak at the M/G_1 transition, when pre-RCs are formed; *de novo* Cdc6p synthesis is required for each round of DNA replication. Transcriptional repression of *CDC6* prevents both pre-RC formation and initiation of DNA synthesis. *cdc6* temperature-sensitive mutants are defective in replication initiation; established pre-RCs are thermolabile in a *cdc6* temperature-sensitive mutant. Cdc6p must be present before late G_1 to permit pre-RC formation. Cdc6p and its homologs also show sequence similarity to subunits of eukaryotic and prokaryotic **clamp loaders** such as replication factor C1 (RFC1), which load ring-shaped DNA polymerase processivity factors onto DNA.

During the early G_1 phase of the cell cycle, the MCM proteins form a ring-shaped heterohexamer that (as a whole) binds as a head-to-head dimer to chromosomal replication origins (Evrin *et al.*, 2009; Remus *et al.*, 2009) and assembles as part of the pre-RC (Figure 5.6). MCM later acts as a replicative helicase and is thus required for cell cycle progression as well as DNA replication initiation and elongation. Pre-RC assembly is also called replication **licensing** of chromosomes prior to DNA synthesis during the S phase. Initially, the MCM complex is loaded at origins in an inactive form, which is then activated during transition to the S phase, in a process that requires both **Cdc7p kinase** and **CDK (cyclin-dependent kinase)** (Bochman and Schwacha, 2009; Remus and Diffley, 2009; Araki, 2010); it appears that the action of both these factors is necessary in all species that have been examined. The Cdc7 kinase phosphorylates the N-terminal tails of Mcm2p, Mcm4p, and Mcm6p, and probably induces a structural change in the MCM complex. Activation of the MCM is associated with the recruitment of many other factors to the

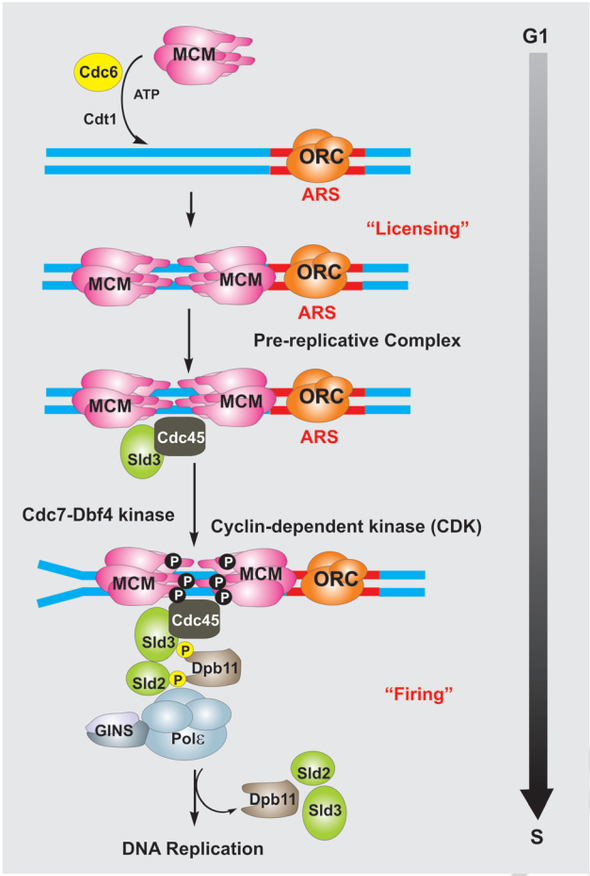


Fig. 5.6 Complex in the initiation of yeast DNA replication. Phosphorylation of residues in Mcm2, 4, and 6 by Cdc7p is indicated in black; phosphorylation sites by Cdk in Sld2p, Sdl3p, and Dbp11p are indicated in yellow.

origin, and initial unwinding of the duplex DNA allows the establishment of two DNA replication forks with opposite polarity. After initiation, replication proceeds bidirectionally away from the origin, until each fork meets another from a neighboring origin, at which point replication of that part of the chromosome is terminated.

The first factor required for the progression of DNA replication forks in addition to MCM was identified as the budding yeast protein **Cdc45p** (Aparicio, Weinstein, and Bell, 1997; Labib, Tercero, and Diffley, 2000). The MCM complex is formed only during the S phase and the assembly process can occur only *in situ* at nascent replication forks (Zou and Stillman, 2000; Masai *et al.*, 2006; Sheu and Stillman, 2006; Im *et al.*, 2009). Cdc45p binds stably to MCM as part of a larger complex of proteins built at replication origins, the existence of which is dependent on an additional component known as the **GINS complex** (named for Go, Ichi, Nii, and San for five, one, two, and three in Japanese), which is composed of four subunits, Sld5p, Psf1p, Psf2p, and Psf3p, distantly related to each other; the GINS complex was first identified in *S. cerevisiae* and is currently the last replication factor conserved in all eukaryotes to be identified (Kubota *et al.*, 2003; Takayama *et al.*, 2003). Studies in yeast first showed that both Cdc7p and CDK are required for the firing of early and later origins of replication throughout the S phase. Cdc7p kinase acts in association with an essential regulatory subunit called Dbf4p (dumbbell-forming 4). CDK also phosphorylates the MCM complex, but the major targets during the initiation of chromosome replication are **Sld2p and Sld3p** (Tanaka *et al.*, 2007; Zegerman and Diffley, 2007), two proteins essential for DNA replication and complex formation with Cdc45p and Dpb11p, which latter is the replication initiation factor that loads DNA polymerase ε onto the complex. The phosphorylated forms of Sld2p and Sld3p appear to be bridged by **Dpb11p**; the N-terminal pair of BRCT (BRCA1 C-terminus) repeats of Dpb11p bind Sld3p, while the C-terminal pair of BRCT repeats bind Sld2p. In yeast it has been established that neither Sld2p nor Sld3p are incorporated into the "replisome." The cell cycle-regulated phosphorylation triggers initiation of DNA replication, including blocking reinitiation in the G_2/M phase. The activation of the two protein kinase complexes, Cdc28/B cyclins and Cdc7p/Dbf4p, serves as the final signal for replication fork movement, whereupon the DNA replication machinery, including DNA polymerases and **PCNA**, initiates DNA synthesis (Figure 5.7). Other S- and

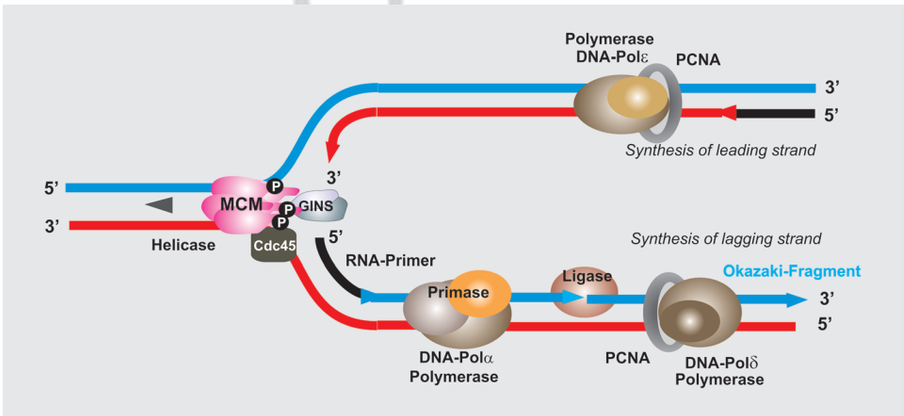


Fig. 5.7 Scheme of DNA replication. Black, RNA; red, leading strand synthesized; blue, lagging strand synthesized.

M-phase Cdks block the rebinding of MCMs to chromatin at ORCs and prevent reinitiation of replication until mitosis is complete.

In yeast, ORC also plays a role in the establishment of silencing at the mating-type loci *HML* and *HMR*. ORC participates in the assembly of transcriptionally silent chromatin at *HML* and *HMR* by recruiting the Sir1p silencing protein to the *HML* and *HMR* silencers.

5.1.3.2 Replication Machinery

Take the next fork. DNA replication is accurately and temporally regulated during the cell cycle in all eukaryotes. After installation of the **replicative fork**, components needed in the DNA- replication machinery, including DNA polymerases, PCNA, and additional factors, initiate DNA synthesis. Similar mechanisms are likely to operate in other organisms, including fission yeast, since homologs of pre-RC proteins and its regulators have been identified in many organisms.

5.1.3.2.1 DNA Polymerases

Yeast contains several multisubunit **DNA polymerases**, denoted by Greek letters α (alpha), γ (gamma), δ (delta), ε (epsilon), η (eta), and ζ (zeta) (*cf.* Table 5.5). Only DNA polymerases α, ε, and δ participate in processive nuclear DNA replication, while DNA polymerase γ is reserved for the replication of mitochondrial DNA; the others (polymerase η and polymerase ζ) serve as repair factors. According to their different tasks, the polymerases are endowed with different enzymatic activities: (i) polymerase α primes the leading and lagging strands, (ii) polymerase ε extends the leading strand, and (iii) polymerase δ extends Okazaki fragments of the lagging strand.

Table 5.5 DNA polymerases, their subunits, and associated factors.

Complex	Subunits	Functions
POL1 α		DNA polymerase α contains four subunits; required for DNA replication; it is involved both in initiation and in priming Okazaki fragments during lagging strand elongation; it has no associated proofreading exonuclease activity; the two smaller subunits form the primase activity that synthesizes short RNA primers in DNA replication; in yeast, DNA polymerase α activity is required for premeiotic DNA replication and sporulation and for DSB repair, but not for other DNA repair synthesis
	Pol1p	*POL1* is an essential gene encoding the largest subunit of DNA polymerase α
	Pol12p	B subunit, executes its essential function during the initiation of DNA replication
	Pri2p	primase synthesizes short RNA primers
	Pri1p	primase synthesizes short RNA primers
	Ctf4p	chromatin-associated protein; required for sister chromatid cohesion. Interacts with DNA Pol1p (Hanna *et al.*, 2001)
	Hcs1p	hexameric α-associated DNA helicase A involved in lagging-strand DNA synthesis; contains single-stranded DNA stimulated ATPase and dATPase activities; RPA stimulates helicase and ATPase activities
POL2 ε		chromosomal DNA replication polymerase that exhibits processivity and proofreading exonuclease activity; also involved in DNA synthesis during DNA repair; interacts extensively with Mrc1p, an S-phase checkpoint protein, that stabilizes Pol2p and Tof1p, phosphorylated by Mec1p; protects uncapped telomeres (Osborn and Elledge, 2003; Tsolou and Lydall, 2007)
	Pol2p	catalytic subunit
	Dpb4p	shared subunit of DNA polymerase ε and of ISW2/yCHRAC chromatin accessibility complex; involved in both chromosomal DNA replication and in inheritance of telomeric silencing
	Dpb2p	second largest subunit; required for normal yeast chromosomal replication; expression peaks at the G_1/S boundary; potential Cdc28p substrate
	Dpb3p	third-largest subunit; required to maintain fidelity of chromosomal replication and for inheritance of telomeric silencing; mRNA abundance peaks at the G_1/S boundary
	Dpb11p	replication initiation protein that loads DNA polymerase ε onto pre-RCs at origins; checkpoint sensor recruited to stalled replication forks by the checkpoint clamp complex where it activates Mec1p
POL3 δ		DNA polymerase δ is involved in many aspects of DNA metabolism, including various types of repair, and both leading and lagging strand elongation; contains three subunits
	Pol3p	catalytic subunit
	Pol31p	second subunit
	Pol32p	third subunit
POL η	Rad30p	belongs to Y-family of DNA polymerases; recruited to stalled replication forks following the monoubiquitination of PCNA by Rad6p–Rad18p in response to DNA damage. polymerase η can replace polymerase δ in the replication holoenzyme
POL ζ	Rev7p	involved in translesion pathway; inefficient DNA polymerase repair across damaged base pairs; cooperates with Pol3p/Rev1p
	Rev3p	
	Rev1p	involved in translesion pathway; member of Y-family; deoxycytidyltransferase
PCNA	Pol30p	homotrimeric ring-shaped complex that encircles DNA and functions as a sliding clamp and processivity factor for replicative DNA polymerases; PCNA is loaded by RFC1 onto primer-template sites of DNA and directs the replication machinery to the replication fork
POL γ	Mip1p	mitochondrial DNA polymerase

5.1.3.2.2 Replication and Replication Factors

Traffic regulation. Figure 5.7 indicates the single steps that can be distinguished in replication, which will always proceed in the 5′ → 3′ direction, implicating that the respective DNA strand serving as a template is copied beginning at its 3′-end. As DNA synthesis requires a short piece of ribonucleotide sequence, a so-called **primer**, to which deoxyribonucleotides can be added subsequently, this is provided by the primase activity of DNA polymerase α, contained in two of its subunits, Pri1p and Pri2p. For the duplication of the **leading strand** (in the direction of the moving replication fork), the synthesis of one particular RNA primer near the origin is sufficient, since the elongation of the leading strand is a continuous process carried out by DNA polymerase ε. Later in replication, the short RNA primer is eliminated by cleavage of the 5′ → 3′ exonuclease activity residing in the large subunit of DNA polymerase α.

Duplication of the **lagging strand** cannot occur in a continuous manner, but requires the synthesis of shorter DNA fragments (away from the moving replication fork), so-called **Okazaki fragments**, which will be ligated (by DNA ligase, Cdc9p) as replication proceeds. Priming of these fragments by short pieces of RNA is established by the low-fidelity DNA polymerase α and synthesis continued by the high-fidelity DNA polymerase δ. (During semiconservative DNA replication in yeast, the lagging strand DNA polymerase δ produces around 100 000 Okazaki fragments.) Also here, the short RNA primers are eliminated by cleavage of the 5′ → 3′ exonuclease activity residing in the large subunit of DNA polymerase α, while the short gaps are filled by DNA polymerase δ. To synthesize a continuous DNA copy of the lagging strand, each of the polymerase α generated RNA–DNA segments is displaced and filled in by polymerase δ.

When polymerase δ arrives at the 5′-end of the downstream Okazaki fragment, it displaces 2–3 nucleotides of the downstream primer at a time and by the action of the 5′ → 3′ exonuclease (5′-flap endonuclease) Fen1p (or Rad27p), these flaps are cleaved, eventually leaving a ligatable nick for DNA ligase I. However, in some cases, longer flaps are generated, which are bound by the single-stranded DNA-binding complex **replication protein A (RPA)** that concomitantly inhibits cleavage by Fen1p, but promotes cleavage by the tripartite replication factor Dna2p (Balakrishnan *et al.*, 2010). Dna2p has both helicase and endonuclease activities, and functions in the removal of long flaps (around 30 bp) bound by the single-stranded DNA-binding protein RPA during the maturation of Okazaki fragments. It is now believed that Pif1p even is responsible for the generation of these long flaps. However, as long flaps will cause chromosomal instability, it appears that Pif1p and polymerase δ together create a long flap, which then requires cleavage by Dna2p.

Although we have described the enzymic activities involved in DNA replication separately, DNA polymerase has to be considered a dimeric (asymmetrical) complex designated the **DNA polymerase holoenzyme**. Additional substantial elements of the replication machinery are as follows.

A helicase activity that disentangles the two parental DNA strands to move the replication fork forward: Hcs1p is a hexameric DNA polymerase α-associated DNA **helicase A** involved in lagging-strand DNA synthesis; it harbors ATPase and dATPase activities stimulated by single-stranded DNA.

Single-stranded DNA-binding proteins, such as the highly conserved **RPA**, are necessary to prevent single-stranded regions to collapse during replication. RPA is a heterotrimer of three essential subunits (Rfa1p and Rfa2p, the binding domains, and Raf3p) that removes secondary structure from single-stranded DNA. Therefore, RPA also plays a key role in other cellular processes dealing with single-stranded DNA intermediates, such as during recombination, transcription, telomere maintenance, and DNA repair. For example, RPA helps loading the telomerase protein Est1p on telomeres, enhances the assembly of Rad51p on presynaptic filaments (*cf.* Section 7.3.1), or specifically targets the damage repair complex (Rad17p/Mec3p/Ddc1p) to 5′-junctions. The response to DNA damage causes phosphorylation of Raf1p and Raf2p by Mec1p and Tel1p.

PCNA (Pol30p) forms a tripartite sliding clamp (in yeast) that binds to the DNA polymerase raising both its processitivity and catalytic capacity, which in yeast can reach about 4000 nucleotides/s. As a consequence of the asymmetrical duplication, PCNA remains tightly bound to the polymerase of the leading strand. Contrary to this, PCNA of the lagging strand has to open, whenever the polymerase has reached the "last" Okazaki fragment – the enzyme dissociates from its template and has to "jump" to the next primer towards the replication fork.

PCNA shows a number of further activities. It is required for the establishment of sister chromatid cohesion, multiple forms of DNA repair, and various postreplication DNA processing reactions, recruiting proteins involved in cell cycle control, NER, MMR, and base excision repair (BER). Channeling PCNA to distinct functional pathways and regulating its activities in DNA replication and postreplication repair, PCNA is subject to differential modification by both ubiquitin and SUMO (Haracska *et al.*, 2004; Andersen *et al.*, 2008) (see below).

The reconstitution of the clamp is catalyzed by "**clamp loader**" or **RFC1**. It should be noted here that *S. cerevisiae* contains four structurally related complexes known as RFC complexes, which are each composed of four small subunits (Rfc2p/Rfc3p/Rfc4p/Rfc5p) plus a large subunit. Depending on the function the RFC has to fulfill, the identity of this large subunit varies. During DNA synthesis, the large subunit is Rfc1p; during sister chromatid cohesion, Ctf18p takes this role. Elg1p is involved in maintaining genomic stability and Rad24p functions as the large subunit in DNA damage check.

5.1.3.2.3 Postreplication Repair and DNA Damage Tolerance

To be or not to be. In the presence of spontaneous DNA damage, living cells have to maintain and complete DNA synthesis or risk replication fork collapse. Since collapsed

replication forks may not be able to restart, this will lead to DSBs or cell death. Perhaps it is more beneficial for the cell to tolerate DNA damage rather than to allow replication fork collapse. The bypass of replication blocks in eukaryotes is effected in distinct ways, which have been most extensively characterized in yeast: one is the so-called **translesion synthesis pathway(TLS)** involving DNA polymerases η (Rad30p) and ζ (Rev1p/Rev3p/Rev7p), both of which operate at the cost of increased mutagenesis; the other is an error-free pathway that induces template switching.

TLS polymerases have been characterized in many eukaryotes (Yang and Woodgate, 2007). Essentially all of them except one (i.e., polymerase ζ) are members of the so-called Y-family polymerases that lack a $3' \rightarrow 5'$ proofreading exonuclease activity and contain relatively nonrestrictive active sites compared with the replicative polymerases. Rev1p, the first characterized eukaryotic Y-family member, is a deoxycytidyltransferase that inserts a dCMP efficiently opposite a template abasic site and is probably responsible for 60–85% of the bypass events. It can also insert dCMP across G or A of the template, although to a lesser extent. Polymerase η is able to correctly incorporate AA opposite *cis-syn* thymine–thymine dimers.

Polymerases η and ζ report on DNA damage by interacting with Mrc1p (mediator of replication checkpoint). Mrc1p is expressed in the S phase to stabilize Pol2p of polymerase ε at stalled replication forks during stress; together with Tof1p (topoisomerase I interacting factor) and Cms3p it forms a pausing complex and is phosphorylated by Mec1p (Alcasabas *et al.*, 2001; Lou *et al.*, 2008); a specialized complex formed from Slx1p and Slx4p contributes to stalling replication by hydrolyzing 5′-branches from duplex DNA. The pausing signal is transferred to Rad53p kinase to induce well-known DNA repair pathways (*cf.* Section 5.1.3.4). Homologs of Tof1p have been identified in fission yeast, *Xenopus*, and mammals.

Finally, two lines of important findings lead to an explanation of how the information of DNA damage is transferred: (i) the aforementioned differential modification of PCNA and (ii) the notion that all eukaryotic Y-family polymerases contain both PCNA-interacting peptide and ubiquitin-binding domains, including ubiquitin-binding motifs (UBMs) or ubiquitin-binding zinc fingers (UBZs). The single ubiquitination site in PCNA is K164, at the same time the major SUMO acceptor site (though an additional minor SUMOylation site is available at K127). PCNA can be mono- and polyubiquitinated following DNA damage (Figure 5.8). Only at stalled replication forks is PCNA monoubiquitinated by Rad6p (E2 enzyme) and Rad18p (E3 ligase), thus signaling potential errors and invoking the translesion synthesis by DNA polymerases η and ζ. Polyubiquitination by the E2/E3 system Mms2p–Ubc13p–Rad5p affects the same K146 and links further ubiquitin residues to K63 of bound ubiquitin; this signal causes stalled replication forks to initiate error-free DNA repair (Hoege *et al.*, 2002). One possibility is that template switching is

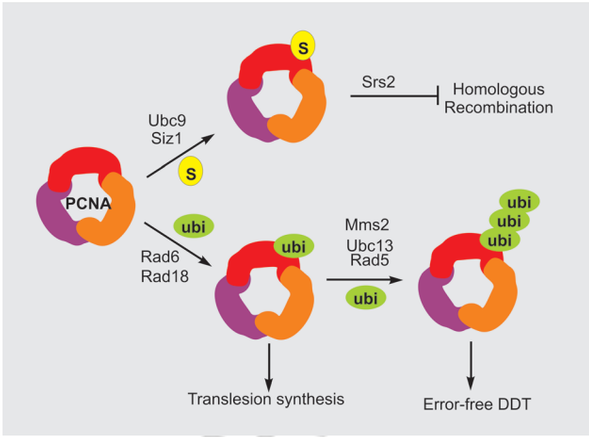


Fig. 5.8 PCNA as a decision maker in DNA repair.

induced, the other is that fork regression occurs. SUMOylation at K146 by yet another E2–E3 complex, Ubc9p–Siz1p, is a third means of modifying PCNA. The known consequence is that SUMOylated PCNA recruits the Srs2p helicase in order to disrupt the Rad51p–single-stranded DNA filament and prevent inappropriate homologous recombination (*cf.* Section 5.1.3.4).

Polymerase α itself has no proofreading capacity and is therefore not involved in DNA postreplication repair. However, it participates in DSB repair.

5.1.3.3 Replication and Chromatin

5.1.3.3.1 Chromatin Reorganization

Plough the way. Histone modification, particularly histone acetylation, has been shown to aid in **disassembly of the nucleosomes** during replication. The MCM helicase complex is associated with a HAT activity that acetylates histones in front of the replication fork (Figure 5.9). After duplication of the DNA, **CAF-I (chromatin assembly factor)** links to PCNA and incorporates newly synthesized H3–H4 dimers with a cytolic acetylation pattern transferred by the histone chaperone Asf1p. Together with PCNA, CAF remains associated with the DNA for up to 20 min after replication. CAF-I consists of three subunits in yeast, Cac1p, Cac2p, and Cac3p, of which Cac3p is similar to proteins found in chromatin-modifying complexes and interacts with SAS HAT. Deletion of any one of the three subunits of yeast CAF results in mild defects in gene silencing at the mating-type loci and the telomeres, a mild sensitivity to UV irradiation, and defects in kinetochore function, in this way reflecting roles in heterochromatin formation, DNA damage repair, and centromere assembly, respectively. Asf1p, *S. cerevisiae* antisilencing factor 1, was first characterized as a factor that, when overexpressed, abolished silencing. Asf1p interacts both physically and genetically with the histone regulator (Hir) proteins to regulate the expression of histones and other proteins. Nucleosome formation is then completed by H2A–H2B delivery.

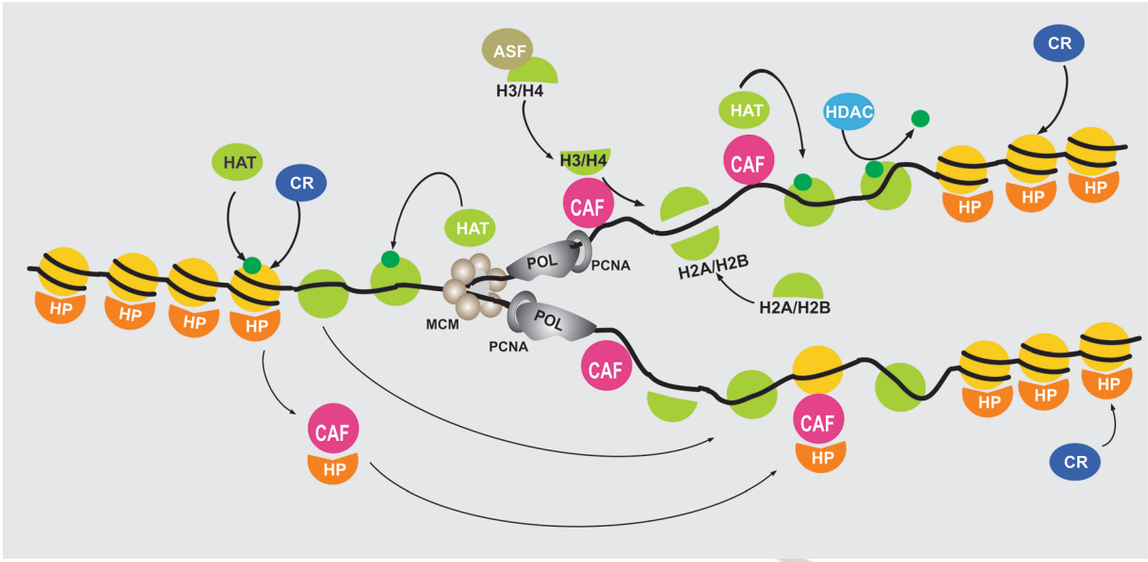


Fig. 5.9 Replication and chromatin. HAT, histone acetyltransferase; HDAC, histone deacetylase; CAF-I, chromatin assembly factor; ASF, antisilencing factor; HP, heterochromatic proteins; CR, chromatin-remodeling factor; dark yellow, normal nucleosomes; light green, modified or disassembled nucleosomes. (After Ehrenhofer-Murray, 2004.)

Chromatin remodeling should help in replication progression by loosening chromatin compaction, thus facilitating the partial disassembly of chromatin before passage of the replication fork. In *S. cerevisiae*, nucleosome positioning by the ORC complex has an important role in the assembly of the pre-RC and facilitates initiation at chromosomal origins of replication (Lipford and Bell, 2001). Also, SWI/SNF was shown to be required for replication initiation. Two chromatin-remodeling complexes of the ISWI type have been implicated in heterochromatin replication, since chromatin remodeling by these complexes may be a prerequisite to move the replication fork through heterochromatin domains. In any case, parental nucleosomes are disassembled into H3–H4 dimers and H2A–H2B dimers during replication, and redistributed randomly onto the two daughter strands. Chromatin is rapidly reconstituted after duplication of the DNA by depositing first H3–H4, then H2A–H2B on the DNA to complete the nucleosomes.

Notably, newly synthesized histones are incorporated along with the parental histones. Thus, the "old" dimers become mixed with "new" H3–H4 and H2A–H2B dimers within the individual new nucleosomes. In order to reproduce the genuine chromatin structure in replicated chromatin, **epigenetic information** has also to be passed on to the progeny. Thus, equalization of the epigenetic patterns of histone modifications between parental and new histones in the replicated chromatin affords removing some modifications and adding others. For example, in *S. cerevisiae*, H4K16 acetylation is a global signal in euchromatic regions, as it prevents the binding of the heterochromatic SIR proteins (*cf.* next Section 5.1.3.3.2) to chromatin outside of their cognate genomic areas. As a consequence, cytoplasmic histones (not acetylated on H4K16) must be acetylated in duplicated euchromatic chromatin. A mechanistic model is that after CAF has deposited newly synthesized H3 and H4, Cac1p subsequently recruits SAS to the chromatin to acetylate H4K16. In a way, SAS takes care of providing a "global" H4K16 acetylation (Osada *et al.*, 2001).

It may well be that the euchromatic pattern is the default pattern after replication and that subsequent steps are required to modify it in different genomic regions. The whole procedure of resetting epigenetic patterns on chromatin needs more than histone acetylation (e.g., re-establishing histone methylation and ubiquitylation patterns after replication). A decisive part in restoring the genuine structure of chromatin after replication is reversion of histone acetylation by the HDACs.

5.1.3.3.2 Silencing and Boundaries

Stop here! Silencing at *HML*, *HMR*, and heterochromatic telomeres in yeast is mediated by the **SIR complex** (silent information regulator), comprised of the two structural proteins Sir3p and Sir4p, as well as Sir2p which is the enzymatic component (deacetylase), and Sir1p. The SIR complex does not bind DNA directly, instead it is recruited to regulatory chromosomal domains bound by Rap1p, Abf1p and the ORC complex (via Sir1p). Unlike repressors that act by binding to promoters, the SIR proteins help repress transcription by creating a silent chromatin structure in a gene- and promoter-independent manner. The recruitment of all these factors leads to assembly of a chromatin-silencing complex and a region of silenced chromatin. Sir1p probably recognizes a silencer element in the DNA through its interactions with ORC, more precisely with Orc1p. Also, a Gal4–Sir1p fusion tethered at *HMR* bypasses the requirement for both the silencer element and ORC in silencing that locus, but still requires passage through

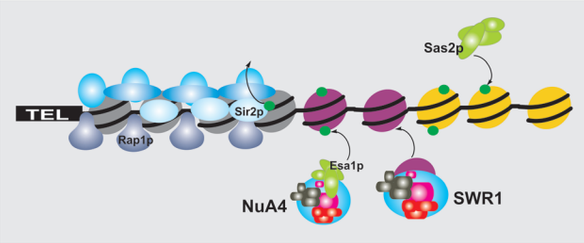


Fig. 5.10 Generation of boundaries near telomeres. Dark green dots, histone acetylation; grey, nucleosomes in heterochromatin; yellow, "normal" nucleosomes; magenta, nucleosomes with histone H2A.Z. (After Pillus, 2008.)

the S phase and the presence of the other SIR proteins for silencing. However, Sir1p is the only SIR protein that appears not to be involved in telomeric silencing.

Once a silencing complex is bound to a nucleosome, Sir2p deacetylates the histone tails of H3 and H4 of the adjacent nucleosome. As the SIR proteins have a higher affinity for H3 and H4 with reduced acetylation, deacetylation creates a binding site for an additional silencing complex. This process repeats until SIR complexes are spread across the entire chromatin region to be silenced.

Sir3p participates in silencing the cryptic mating-type loci and is a key player in maintaining a repressed chromatin structure near telomeres. It appears that Sir3p is recruited by Rap1p to telomeres, where it acts along with Sir2p and Sir4p to maintain silencing. The silenced domains located next to chromosome ends spread inward from the telomeres in proportion to Sir3p levels in the cell and it appears that Sir3p is a structural component of the heterochromatin, as it is detected spreading inward along with the silenced domains. Sir4p seems to act in the maintenance rather than the initiation of silencing at the mating-type loci. Genetic and physical interactions between Sir2p and Sir4p, Sir3p and Sir4p, and Rap1p and Sir4p have been described.

A model for the generation of **boundaries** between heterochromatic (in lila) and euchromatic domains is shown in Figure 5.10. At subtelomeric regions, the acetylase Sas2p of the MYST SAS complex acetylates H4K16, whereby the SWR complex triggers the assembly of H2A.Z-containing nucleosomes. It may be noted here that the H2A variant H2A.Z (Htz1p in *S. cerevisiae*) is estimated to be present in 5–10% of all nucleosomes. Exchange of H2A for H2A.Z in chromatin was discovered to be triggered by the chromatin-remodeling complex SWR1 (Kobor *et al.*, 2004; Krogan *et al.*, 2003; Mizuguchi *et al.*, 2004). The current model is that also perhaps other histone variants are not directly incorporated into chromatin at replication, but that specialized remodeling complexes are deployed after replication-coupled chromatin assembly to swap the conventional histones for the histone variants. After incorporation, H2A.Z is acetylated by Esa1p, the acetylase of the NuA4 complex, which has four of the same subunits that are also present in SWR. These events facilitate the formation of a boundary between heterochromatin (represented by the dark nucleosomes in Figure 5.10) versus euchromatin (represented by the yellow nucleosomes in Figure 5.10), likewise preventing the spread of silencing proteins such as the SIR complex (components in blue in Figure 5.10). Marked is the Sir2p deacetylase (light blue in Figure 5.10), which deacetylates the histones within the heterochromatic region. Methylation of H3 by Dot1p and Set1p (not shown in Figure 5.10) is also important in boundary formation. Recent studies have demonstrated that histone H2A.Z and Set1p act in concert, not only in subtelomeric regions but also throughout the genome, to limit the spread of silent chromatin.

As indicated, Sir2p together with Sir3p and Sir4p is required for all forms of silencing: at telomeres (contributing to the stability and maintenance of telomeric repeats), at the mating-type loci, and at rRNA genes. Surprisingly, the Sir2p sites in most of the yeast chromosomes are accompanied by chromosomal domains maintained in a hypoacetylated state, the so-called **HAST (Hda1p-affected subtelomeric) domains** (Figure 5.11). They appear to represent sort of a facultative heterochromatin because several genes located in this region can be switched on or off depending on growth or developmental signals. In this context, it is interesting to note that a global (as opposed to promoter-targeted) mode of histone deacetylation has been described in *S. cerevisiae* (Robyr *et al.*, 2002), whereby the different HDACs (e.g., Rpd3p and Hda1p) are dedicated to individual genomic territories (Figure 5.11). As an example, global deacetylation by Hda1p is found concentrated to contiguous subtelomeric domains and to regions in the vicinity of centromeres. One possibility is that the HDACs are brought to these regions via chromatin assembly during the S phase. Alternatively, they may act constantly (i.e., throughout the cell cycle). Support for this

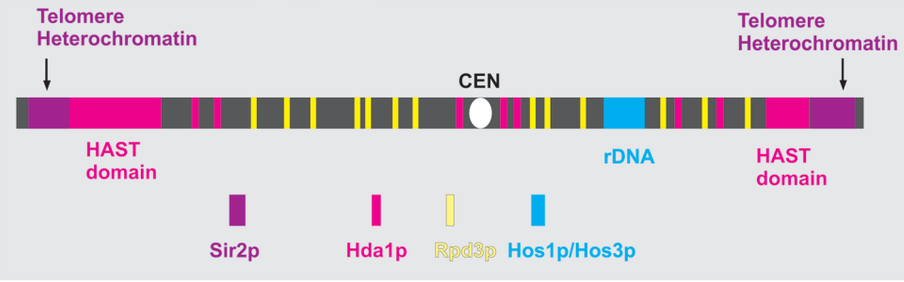


Fig. 5.11 Division of labor among HDACs on an idealized yeast chromosome. The colored blocks indicate domains in which preferred HDACs become active. (After Robyr *et al.*, 2002.)

notion comes from the observation that histone acetylation targeted by a transcriptional activator is rapidly reversed upon removal of the activator and independently of the cell cycle, suggesting that there is a constant equilibrium between acetylation and deacetylation activities in the yeast genome.

At **rRNA genes**, Sir2p is associated with two other proteins in a complex called **RENT**, a regulator of nucleolar silencing and telophase exit: Net1p, which tethers the SIR complex to rDNA, and Cdc14p, a protein phosphatase involved in cell cycle control. RENT is recruited to rDNA by interaction with Fob1p and RNA polymerase I; its function is to suppress mitotic and meiotic recombination between rDNA clusters. Silencing at rDNA mediated by Sir2p appears to prevent or delay the formation of extrachromosomal rDNA circles, which have been shown to segregate to yeast mother cells and promote senescence. Increased dosage of the *SIR2* gene resulted in elevated lifespan, whereas deletion of *SIR2* shortens **lifespan**. In addition, Sir2p has been implicated in cell cycle progression and chromosome stability. Hst2p, a homolog of Sir2p, is localized to the cytoplasm. Overexpression of this enzyme influences nuclear silencing events by derepressing subtelomeric silencing and increasing repression in the rDNA. The general view drawn from these findings is that the metabolic energy status (e.g., NAD^+: NADH ratio) is secured through histone or specific protein deacetylation by Sir2 enzymes requiring NAD^+, in this way preventing unwanted activity.

Sir4p and Sir3p also seem to have a role in the aging of yeast cells. An allele of *SIR4* was found that extends the lifespan of yeast. In strains with this allele, Sir3p and Sir4p are redirected to the nucleolus rather than telomeres. Mutations in *SIR4* that lead to a longer lifespan also result in enhanced rDNA silencing. It may be that the lengthening of lifespan is due to the prevention of formation of extrachromosomal rDNA circles that form through homologous recombination within rDNA arrays, which is inhibited when the rDNA is silenced. For Sir3p, the proportion of components found at telomeres versus the nucleolus decreases as cells age and mutations in *SGS1* or *RAD52* that shorten the lifespan of yeast also result in the redistribution of Sir3p from telomeres to the nucleolus.

The formation of chromatin at centromeres during replication entails the deposition of Cse4p and restructuring centromeric and centromere-flanking nucleosome architecture for the accurate transmission of chromosomes (Hsu *et al.*, 2003). The deposition of Cse4p does not require CAF, but a cooperation of CAF with the Hir proteins will restrict Cse4p to its centromeric location (Sharp *et al.*, 2002). For restructuring, the chromatin remodeler RSC is required. Thus, both CAF-I/Hir and RSC may function in postrecruitment assembly or maintenance of centromeric chromatin.

It is obvious that the dynamic aspects of replication are of outstanding importance and need to be connected to the events governed by the cell cycle. This is covered in more detail in Chapter 7. We still have to deal with the handling of DNA damage during replication. An excellent overview connecting DNA replication to damage checkpoints and cell cycle controls, and at the same time comparing these issues in budding and fission yeasts, is available from Murakami and Nurse (2000).

It might also be interesting to point out here that several of the proteins involved in replication have been found to belong to the novel class of the AAA^+-ATPases as established by sequence comparisons. Likewise, of importance is the participation of components of the ubiquitin–proteasome pathway in the process of DNA replication. Details are discussed in Section 6.2.7.

5.1.3.4 DNA Damage Checkpoints

5.1.3.4.1 Checkpoints During Replication

Tell it to the boss! DNA damage checkpoints are regulatory signal transduction cascades that are triggered by incompletely replicated or damaged chromosomes that provoke cell cycle arrest and DNA repair. Five proteins, Rad17p, Mec1p, Ddc1p, Rad24p, and Mec3p, are required for both the DNA replication and DNA damage checkpoints in budding yeast (Figure 5.11). These proteins are called checkpoint Rad proteins.

Rad17p, Mec3p, and Ddc1p form a **PCNA-like clamp** complex, whereby Ddc1p phosphorylation is dependent on Mec1p. This complex is loaded on to partial duplex DNA, whereby Rad24p becomes a subunit of the corresponding clamp loader, Rad24–RFC, whose other constituents are Rfc2p, Rfc3p, Rfc4p, and Rfc5p. (In other organisms this clamp loader complex is also known as Rad17–RFC and the human equivalent of Rad17p/Mec3p/Ddc1p is called the 9-1-1 complex.) The ATP-binding activity of Rad24p is necessary for the ATPase and clamp-loading activities of the RFC complex. Rad24–RFC interaction with DNA during clamp loading also requires interactions with RPA. Rad24p is phosphorylated by the checkpoint kinase Mec1p.

The Rad24–RFC complex is also operative in homologous recombination during pachytene formation in meiosis (*cf.* Section 7.2.5), as well as in mitotic checkpoints for repair of DSBs. Further, Rad24p is involved in processing DSB ends and recombination partner choice, efficient inducible NER and nonhomologous end-joining (NHEJ), and telomere maintenance through stimulation of Ty1 transposition. Cells lacking Rad24p function are impaired at the various DNA damage checkpoints, show decreased stability at CAG expansion sites, produce nonviable spores, and are delayed entering into meiosis I.

The RAD24 pathway is one of two DNA damage checkpoint pathways, the other involves the RAD9 epistasis group (Mec1p and Rad53p,) that converges on Rad53p phosphorylation.

Mec1p is a member of the evolutionarily conserved subfamily of phosphatidylinositol-3-kinases (*cf.* Section 3.4.3.1), which includes yeast Tel1p. Chk1p is a protein kinase that

provides a link between the checkpoint Rad proteins and the machinery that controls mitosis; it is phosphorylated by Rad53p.

Rad53p is a kinase that is required for both the DNA replication and DNA damage checkpoints; it has two FHA (forkhead associated) domains. Rad53p will target the transcription factor Swi6p as well as the kinases Dun1p and Dbf4p. The activation of the two kinases induces the G_1 cyclins and the ribonucleotide reductase; firing of late replication origins is inhibited. The activation of Rad53p follows two principles: autophosphorylation and phosphorylation by Mec1p, assisted through binding of Rad9p. The N-terminal FHA domain of Rad53p interacts with type 2C phosphatases, Ptc2p and Ptc3p, which are required for inhibition of Rad53p function, when Rad53p is mainly involved in the DNA replication check.

Loss of Rad53p in yeast leads to multiple defects, including impaired checkpoint activation, inability to recover from replication blocks, X-ray sensitivity, and excess histone accumulation resulting in slow growth and chromosome loss. Rad53p is the homolog of *S. pombe* Cds1 and human CHK2. Mutations in the human tumor suppressor CHK2 have been associated with sporadic cancer as well as familial breast cancer and Li–Fraumeni syndrome

Rad9p preferably binds to the C-terminal FHA domain of Rad53p, demonstrating that this association is required for transduction of the DNA damage checkpoint signal. By mediating phosphorylation of important effector kinases, Rad9p facilitates the amplification of initial signals in response to DNA damage. In view of the ability of Rad9p to associate with DSBs (through a so-called Tudor domain), it is believed that Rad9p even induces the checkpoint signal transduction cascade by acting as a **DNA damage sensor**. Rad9p is required throughout the cell cycle; it has been shown to function from G_1/S, through S, up to G_2/M. During normal progression of the cell cycle, Rad9p is phosphorylated, but becomes hyperphosphorylated by Mec1p (and Tel1p) in response to DNA damage. Activated Rad9p then stimulates Mec1p phosphorylation of the effector kinases Chk1p and Rad53p. Rad9p contains two BRCT domains in its C-terminus that facilitate Rad9p interactions after DNA damage. Rad9p purifies in two distinct complexes – the larger 850 kDa complex contains Rad9p and the chaperones Ssa1p and Ssa2p; the smaller 560 kDa complex additionally includes Rad53p.

Chk1p and Rad53p phosphorylation mediate a number of processes associated with cellular arrest, such as transcriptional upregulation of DNA damage repair genes, transcriptional repression of the cyclins, and stabilization of replication forks. Chk1p and Rad53p communicate to the **mitotic apparatus** through Pds1p and Cdc5p (Figure 5.12), but in a different way.

Pds1p (securin) is an anaphase inhibitor that is required for the DNA damage and spindle checkpoints (see Section 7.2.2.2). Pds1p is phosphorylated in response to DNA damage – a phosphorylation that is dependent on budding yeast Chk1p and Mec1p, but not Rad53p. Furthermore, yeast

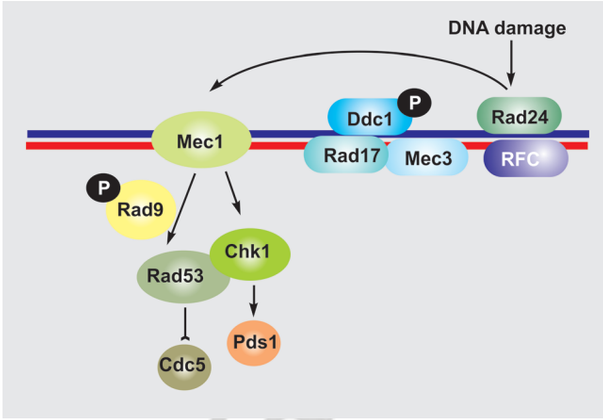


Fig. 5.12 DNA damage checkpoints in *S. cerevisiae*.

Chk1p phosphorylation depends on Mec1p, and the Chk1p protein binds and phosphorylates Pds1p. These findings implicate that Pds1p acts downstream of Chk1p and that Chk1p acts downstream of Mec1p.

Budding yeast **Cdc5p**, a member of the polo-like family of kinases, mainly involved in the exit from mitosis, is blocked at DNA damage by Rad53p, meaning that this cell cycle arrest delays chromosome segregation until the damage has been corrected (see Section 7.2.2.3).

5.1.3.4.2 DSB Repair

Where there's a will there's a way. A further DNA damage checkpoint is effective in the repair of **DSBs**. This pathway involves a number of Rad proteins that induce strand exchange at DSBs as well as DSB repair proteins; the pathway functions both at vegetative growth and at meiosis. Rad51p, Rad52p, Rad54p, Rad55p, and Rad57p are interacting factors. Rad58p (or Mre11p) forms a complex (the **MRX complex**) with Rad50p and Xrs2p as partners, which functions in DSB repair and stabilization of telomeres. The complex is stable at a stoichiometry of 2 : 2 : 1 (Mre11/Rad50/Xrs2), operates in DNA binding and unwinding, and possesses endo- and exonuclease activity. Further, the MRX complex facilitates DSB repair by NHEJ (see below) as well as the introduction of DSBs, which are obligate in meiosis. The MRX complex is conserved from Archaea to humans. While the Mre11p and Rad50p components are highly conserved, Xrs2p is only weakly and only represented in eukaryotes. Null mutants in yeast are viable, but extinction of one of the three components in vertebrates causes embryonic lethality or cell death. The complex Rad59p–Rad52p anneals to complementary single-stranded DNA.

Cells have developed two distinct mechanisms for the repair of DSBs (Prakash *et al.*, 2009). The **NHEJ** pathway is particularly important during the G_1 and early S phases of the cell cycle, and repair by this pathway usually involves only a limited amount of DNA end-processing. Briefly, the following pathway in NHEJ can be envisaged. Dnl4p, a specialized DNA ligase (known as DNA ligase IV in mammals)

assembles together with other parts of the NHEJ machinery at the DSB sites. Dnl4p interacts with Lif1p – a reaction that stabilizes Dnl4p and promotes its activity. DNA polymerase IV (gene *POL4*) then undergoes pair-wise interactions with Dnl4p–Lif1p and Rad27p to effect repair of the DNA DSB. Further components mediating this process are Ydl012c, a tail-anchored plasma protein, also involved in stress response; Doa1p, forming a complex with Cdc48p, that promotes efficient NHEJ; and, most importantly, the Ku heterodimer, Yuk70p and Yuk80p, which otherwise is known to be involved in telomere maintenance. The central β-barrel ring structure in Ku binds DNA by slipping the DSB through this ring. The C-terminus of Yku80p is oriented toward the DSB and provides contact with Dnl4p, while the C-terminus of Yku70p is positioned away from the DSB end. A further component associated with the NHEJ machinery is Lrp1p – a nucleic acid-binding protein of the nuclear exosome. The great disadvantage of this repair pathway is its susceptibility to incorporation of wrong nucleotides, thus inducing mutations.

By contrast, **homologous recombination** typically utilizes the intact sister chromatid to guide the repair process, and it is active mostly during the S and G_2 phases of the cell cycle. For this type of recombination it is required that 3′-single-stranded DNA tails are generated by extensive processing of the DSB ends (Figure 5.13). The DSB repair pathways also differ in the extent to which they are conservative. NHEJ suffers not only from the incorporation of wrong nucleotides, but also from the gain or loss of nucleotides and is thus error-prone. However, when the sister chromatid is used as the information donor, homologous recombination is largely an error-free means of repair.

After DSB formation, nucleolytic processing of the ends results in a pair of 3′-single-stranded DNA tails, which recruit the **recombinase protein Rad51p**, leading to the assembly of an extended, right-handed helical "**Rad51 filament**," commonly referred to as the presynaptic filament. This will find a homologous DNA sequence with the assistance of one of the accessory factors, such as the Swi2/Snf2-related DNA motor protein Rad54p or Rdh54p (both DNA-dependent DNA ATPases stimulating strand exchange), and invade it to form a displacement loop ("D-loop"). This D-loop can be resolved in different ways, each having different consequences. In the canonical DSB repair pathway, DNA synthesis is initiated from the primer terminus of the D-loop in order to enlarge the structure. This procedure allows capture of the second end of the break, resulting in a DNA intermediate that generates a double Holliday junction (dHJ). This junction can be cleaved by a specialized endonuclease called HJ resolvase, to yield a mixture of cross-over and non-cross-over products (Colavito, Prakash, and Sung, 2010). Alternatively, the "synthesis-dependent strand annealing pathway" utilizes a specialized DNA helicase system (Sgs1p, Top3p, Rim1p) that can resolve the D-loop structure to generate exclusively non-cross-over recombinants. Sgs1p is a helicase of the RecQ family that forms a complex with Top3p (topoisomerase III) and Rmi1p to relax single-stranded negatively supercoiled DNA. A third possibility is the interference of Srs2p, which through phosphorylation by Cdk is able of aborting homologous recombination by disrupting the Rad51p presynaptic filament. Subsequently, the D-loop is unwound by the helicase Mph1p, leading to entirely non-cross-over products. Each of these factors has a human counterpart.

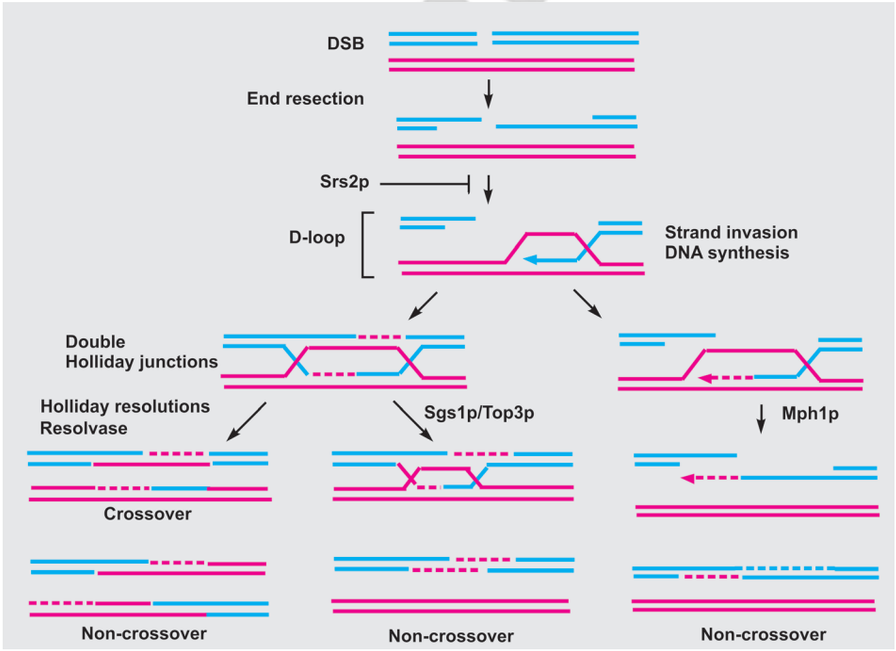


Fig. 5.13 DSB repair. (After Colavito, Prakash, and Sung, 2010.)

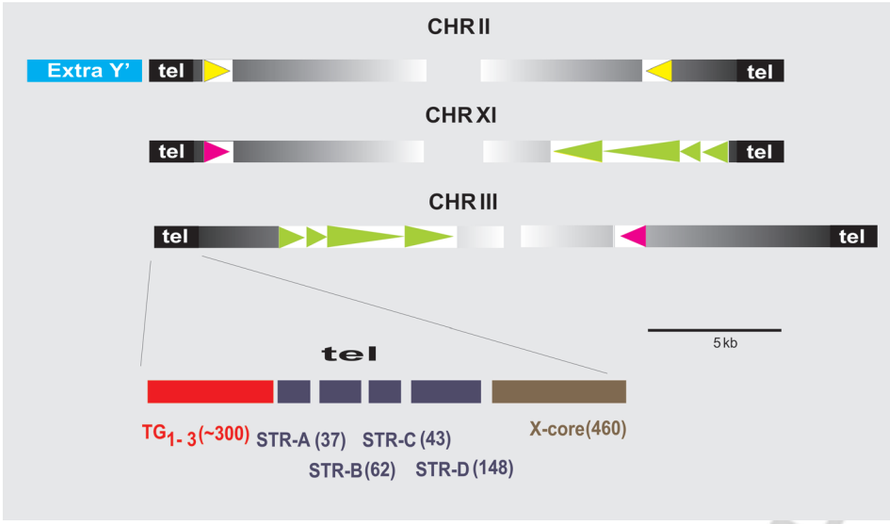


Fig. 5.14 Structures of some yeast telomeres. Colored arrows indicate repeated subtelomeric gene sequences.

5.1.4 Telomeres

Sealing the chromosome ends. Telomeres are specialized DNA sequences that enable complete **replication of chromosome ends** and prevent their degradation in the cell (Figure 5.14). The first yeast telomeres were cloned in linear plasmid vectors by Szostak and Blackburn (1982). Remarkably, both these researchers together with C.W. Greider received the Nobel Prize for Physiology or Medicine in 2009 for their "discovery that chromosomes are protected by telomeres and the enzyme telomerase."

Later, analysis of telomeric sequences by Louis and coworkers in conjunction with the yeast genome sequencing project revealed that all yeast chromosomes share characteristic telomeric and subtelomeric structures (Louis, 1995; Louis and Borts, 1995). Telomeric ($TG_{1-3}/C_{1-3}A$) repeats, some 300 nucleotides in length, are found at all telomere ends. Thirty-one of 32 of the yeast chromosome ends contain the X core **subtelomeric elements** (400 bp) and 21 of 32 of the chromosome ends carry an additional Y′ element. There are two Y′ classes, 5.2 and 6.7 kb in length, both of which include an open reading frame (ORF) for an RNA helicase, Pif1p, which is a catalytic inhibitor of telomerase in yeast (Zhou *et al.*, 2000). **Y′ elements** show a high degree of conservation, but vary among different strains (Louis and Haber, 1992). Experiments with the *est1* (ever shorter telomeres) mutants, in which telomeric repeats are progressively lost, have shown that the senescence of these mutants can be rescued by a dramatic proliferation of Y′ elements (Lundblad and Blackburn, 1993). Several additional functions have been suggested for these elements, such as extension of telomere-induced heterochromatin or protection of nearby unique sequences from its effects and a role in the positioning of chromosomes within the nucleus (Palladino and Gasser, 1994; Zakian, 1996a; Zakian, 1996b).

How to end the "open end." Because of their "open-end" structure, telomeres have to be replicated by a specialized **telomerase system** (RNP complex that is essential for maintenance of telomeres) (Cohn and Blackburn, 1995). Yeast telomerase is a reverse transcriptase that elongates the single-stranded G-rich 3′-protruding ends of chromosomal DNA using an RNA molecule that is part of the telomerase complex. The extended strand provides a template for synthesis of the lagging strand by DNA polymerase, thus preventing the otherwise inevitable loss of terminal DNA at each round of replication. Telomerase consists of the gene products of three *EST* genes (Est1p, Est2p, and Est3p) (Taggart, Teng, and Zakian, 2002; Lundblad, 2003; Taggart and Zakian, 2003) (whereby Est2p acts as the catalytic subunit) as well as an RNA component (TLC1) that is employed as a template in the synthesis of telomeric DNA (Brigati *et al.*, 1993) (Figure 5.15), and Cdc13p, an essential multifunctional and single-stranded DNA-binding protein, whose main function is telomere capping (Dubrana, Perrod, and Gasser, 2001). Without functional Cdc13p yeast is not viable.

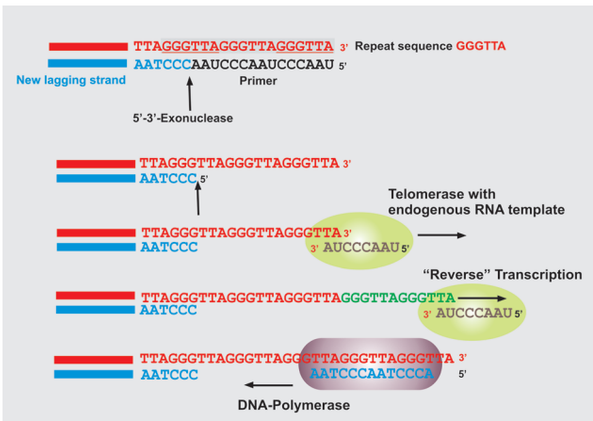


Fig. 5.15 Scheme of telomere replication.

Telomere replication machinery. Mutations in any of these five genes lead to progressive telomere shortening, the so-called ever shorter telomeres (*est*) phenotype, followed by cell death. *CDC13* is the only essential gene among the *EST* genes. Est2p and TLC1 RNA form the core of telomerase, while Est1p, Est3p, and Cdc13p, which are dispensable for *in vitro* telomerase catalytic activity, play regulatory roles. Cdc13p, a single-stranded DNA-binding protein required for telomere maintenance and elongation, binds to Est1p and this interaction is necessary for recruiting telomerase to the chromosomal ends. Est1p, Est2p, and Est3p all bind to the TLC1 RNA template, and Est1p also binds to 3′-ends of single-stranded DNA. Est1p forms a stable complex with TLC1 in the absence of Est2p or Est3p, while association of Est3p with the enzyme requires an intact catalytic core. Est1p and Est3p are stable components of the telomerase holoenzyme.

In addition, **telomere replication** depends on

i) The **TRF1 complex**, consisting of Ku70 (Yku70p/Hdf1p) and Ku80 (Yku80p/Hdf2p) proteins and interacting with Cdc13p, which is also crucial for non-homologous DNA DSB repair and protects telomeres against nucleases and recombinases (Stellwagen *et al.*, 2003; Fisher, Taggart, and Zakian, 2004) (see above). The N-terminal region of Cdc13p is crucial for the protein's activity due to its various interactions with different binding proteins. The telomere-capping function is mediated through its interaction with Stn1p and Ten1p – essential proteins required for telomere length regulation. Cdc13p plays a role in telomere replication through its interactions with the Pol1p catalytic subunit of DNA polymerase α and an essential subunit of telomerase, Est1p. Cdc13p and Est1p also recruit and activate the telomere-bound Est2p catalytic subunit of telomerase for its replication. The telomerase recruitment step is regulated by the yeast Ku heterodimer (Yku70p–Yku80p) and Stn1p, which impart positive and negative control on the Cdc13p–Est1p interaction. Cdc13p is regulated by the phosphorylation of the SQ/TQ motif in the telomerase recruitment domain by the checkpoint kinases, Mec1p and Tel1p. Mutation in Cdc13p results in abnormal uncapped telomeres with long exposed G-strands leading to activation of the RAD9 DNA damage pathway, cell cycle arrest at the G_2/M phase and cell death.

ii) A number of **RAD proteins** (Rad50p, Rad51p, and Rad52p), which are involved in stabilizing the telomeres but also participate in homologous recombination and DSB repair (*cf.* Section 6.5.2.2.4).

iii) **Sgs1p**, a helicase, preventing deleterious recombination between telomeric sequences (*cf.* Section 6.5.2.2.4).

iv) A number of other proteins, such as the helicase Pif1p detected in this function by Zakian's group (Zhou *et al.*, 2000). The participation of helicase Pif1p in telomere replication (Boule and Zakian, 2006) as well as the involvement of the telomere replication apparatus in healing DNA breaks (Bianchi, Negrini, and Shore, 2004) has been resolved in the budding yeast model (Figure 5.16). As mentioned above (Section 5.1.3.2.2), Pif1p also helps flap elongation during Okazaki fragment maturation, while Rrm3p, a helicase belonging to the Pif family, appears to be involved in replication fork progression (Boule and Zakian, 2006).

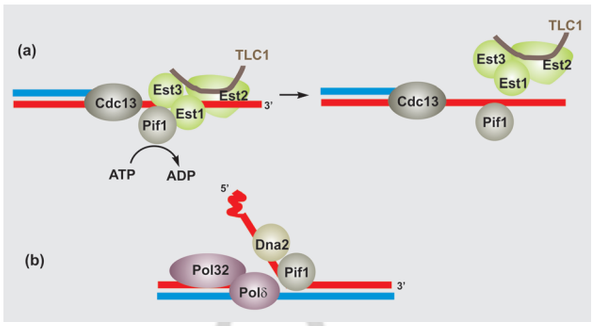


Fig. 5.16 Models explaining Pif1p action at yeast telomeres (a) and during Okazaki fragment maturation (b). (After Boule and Zakian, 2006.)

The secondary structure of the **TLC1 RNA** gave a clue to the RNA–protein interactions that are necessary for the assembly and activity of the telomerase complex. A base-paired domain immediately adjacent to the template acts as a template boundary to terminate each cycle of reverse transcription, while three stem–loop structures function as protein-binding sites for the Est2p and Est1p telomerase subunits, and for the Ku heterodimer (Yku70p–Yku80p). Binding of Est1p with TLC1 is thought to build a bridge between the catalytic Est2p and the telomere-bound Cdc13p. Interaction of TLC1 RNA with the Ku dimer promotes the addition of telomeres to broken chromosome ends, whereby damaged DNA is repaired by capping the broken end with telomeric DNA. Further, TLC1 near its 3′-end binds to the heteroheptameric Sm ring complex, which is also found in many of the spliceosomal small nuclear RNPs (snRNPs; *cf.* Section 4.2). It has been postulated that the Sm proteins play a role in the intracellular transport, assembly and maturation of such RNP complexes.

Protection of telomeres and telomere length. Several factors have been found to be implicated in **stabilizing telomeres**; one of the earliest identified was Rap1p (repressor-activator protein), binding to both silencer and activator elements (Shore and Nasmyth, 1987; Shore *et al.*, 1987; Kurtz and Shore, 1991). David Shore and colleagues also found two proteins, Rif1p and Rif2p (Hardy, Sussel, and Shore, 1992; Wotton and Shore, 1997), and a SIR complex (Sir2p, Sir3p, and Sir4p) (Moretti *et al.*, 1994) interacting with Rap1. Remarkably, these factors are involved in telomere length regulation (Lustig, Kurtz, and Shore, 1990; Hardy, Sussel, and Shore, 1992; Wotton and Shore, 1997; Shore, 2001; Del Vescovo *et al.*, 2004).

In telomerase-deficient yeast cells, **telomeres shorten** progressively (in about 60 generations), leading to a shortening of telomeres and increased senescence. Two types of survival pathways are known to be induced upon defects in the telomerase system, which consist of telomere elongation by break-induced replication: (BIR) type I survivors maintain short $TG_{1–3}$ repeats, but amplify the Y′ repeats, while type II survivors amplify the $TG_{1–3}$ repeats to several kilobases in length, but do not amplify the Y′ elements (Lydall, 2003). To date, the data providing an answer to the question of how telomeres are distinguished from DSBs during their preparation for telomerase elongation come mainly from *S. cerevisiae*, in which the early events in telomerase lengthening are indistinguishable from what occurs at DSBs (Sabourin and Zakian, 2008). During chromosome replication in yeast, telomeres connect to the SPBs and there are multiple pathways for telomere tethering (Taddei and Gasser, 2004).

The first mutations found to affect **telomere length**, *tel1* and *tel2*, were identified when a collection of *S. cerevisiae* mutants was screened by Southern hybridization for strains with short telomeres (Lustig and Petes, 1986). When the yeast *TEL1* gene was sequenced (Greenwell *et al.*, 1995), its closest homolog in the database was the human gene for the checkpoint protein kinase ATM (ataxia telangiectasia mutated). Tel1p, primarily involved in telomere length regulation, also contributes to cell cycle checkpoint control in response to DNA damage; it is functionally redundant with Mec1p. A relatively recent review focuses on findings that shed light on the role of the ATM-like kinase and other checkpoint and repair proteins in telomere maintenance, replication, and checkpoint signaling (Sabourin and Zakian, 2008). *TEL2* turned out to encode an essential DNA-binding protein required for telomere length regulation and the telomere position effect in yeast (Runge and Zakian, 1996).

Telomere length regulation is an issue long discussed as a decisive phenomenon in cellular senescence and aging (Shore, 1997; Smeal and Guarente, 1997; Shore, 1998; Blackburn, Greider, and Szostak, 2006) pertinent to all eukaryotic organisms.

5.1.5 Transposons in Yeast

5.1.5.1 Classes of Transposable Elements

Unwanted genetic invaders? Several types of transposons have been classified: transposons, which encode a transposase required for transposition (class I), and retrotransposons, which use a retrotranscriptase encoded in their genome for retrotransposition (class II). Transposons are found in a large variety of eukaryotes, and often both types and different subfamilies of transposons are represented in a particular organism. With respect to gene organization and expression strategies, the retrotransposons are highly related to animal retroviruses.

In *S. cerevisiae*, only several types of class II retrotransposons (or retroposons) have been identified. In other yeasts, however, class I elements have also been characterized (*cf.* Chapter 15). The unique properties of the yeast retroposons occupying some 3% of the genome have attracted the interest of several researchers. The relationships with retroviruses have made these elements useful model systems to study their gene structures and functions, their replication, and the interactions between them and their host. The final organization of these elements only became apparent with the deciphering of the whole genome sequence (*cf.* Section 12.2.)

5.1.5.2 Retrotransposons in *S. cerevisiae*

Soon after the detection of "repeated gene families" in *Drosophila* (Rubin, Finnegan, and Hogness, 1976; Finnegan *et al.*, 1978), evidence for transposition of dispersed repetitive DNA families was found in yeast in the laboratory of Ron Davis (Cameron, Loh, and Davis, 1979). These transposable elements, about 6 kb in length, were named Ty (transposons yeast) elements and shown to be associated with DNA rearrangements in studies that were mainly guided in the laboratory of Gerald Fink and his collaborators (Roeder and Fink, 1980; Roeder *et al.*, 1980; Fink *et al.*, 1981). Physical analysis of chromosomal regions harboring Ty1 elements revealed that they could cause deletions in nearby genes (Silverman and Fink, 1984).

5.1.5.2.1 Ty Elements and their Genomes The Ty elements belong to a ubiquitous group of retrotransposons containing long terminal repeats (LTR) at both extremities of the element. Different types of such elements exist in eukaryotes as diverse as insects, plants, fungi, yeasts, and fishes. Recently, fossils of LTR retrotransposons were identified in mammals at a very low copy number. The structure of LTR retrotransposons is comparable to that of retroviruses that replicate via an mRNA intermediate (Boeke *et al.*, 1985). Commonly, two genes are found in LTR retrotransposons, representing the homologs of the retroviral *gag* and *pol* genes. The *gag* gene of retroviruses encodes structural proteins of the virus particle and the retroviral *pol* locus encodes a polyprotein with protease (PR or *prot*), integrase (IN or *int*), reverse transcriptase (RT), and RNase H (RH) catalytic domains. Arrangement and functions of these entities in LTR retrotransposons largely correspond to those in retroviruses.

Some elements, such as *gypsy* from *Drosophila melanogaster*, harbor a third gene homologous to the retroviral *env* gene encoding a protein similar to the envelope of infectious viral particles (Rubin, Finnegan, and Hogness, 1976; Finnegan *et al.*, 1978). It is noteworthy that the existence of virus-like particles (VLPs), which constitute the transposition-competent structures, have been shown for the yeast retroelements, Ty1 and Ty3. However, the VLPs cannot be transmitted horizontally and are thus not infectious to neighboring yeast cells.

LTR retrotransposons have been divided into two distinct groups on the basis of sequence similarities of their reverse transcriptases and organization of the subunits within their

Table 5.6 Retrotransposons in *S. cerevisiae* αS288C

Chromosome	Ty1	Ty2	Ty3	Ty4	Ty5	Total
I	A[a] (1)					1
II	BL[d], BR[a] (2)	B (1)				3
III		C[a] (1)			C (1)	2
IV	DR1, DR2[b, a], 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[d] (1)			6
VIII	H[c] (1)			H (1)		2
IX			I (1)			1
X	JR1, JR2 (2)			J[a] (1)		3
XI	none					0
XII	LR1, LR2, LR3, LR4[a] (4)	LR1, LR2[a, b] (2)				6
XIII	ML1[b], ML2, MR1[d], MR2[b] (4)					4
XIV	NL1[a], NL2[d] (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	2[d]	1	2	1		
Solo elements or remnants						268

Nomenclature of the Ty elements is as in the conventional sequence annotations; (e.g., ML1 is the first element on the left arm of chromosome XIII). Number of elements on a given chromosome is in brackets.

a) Sequence ambiguities versus consensus.
b) Nonintact element.
c) Insert of 114 bp corresponding to 38 amino acids (cloning artifact?).
d) Ty1_BL, Ty1_DR2, Ty1_MR1, and Ty1_NL2 belong to a second subtype of Ty1 elements having a variant TYA protein sequence.

pol genes, which in the ***copia*** group are arranged in the order PR, IN, RT, and RH, but in the ***gypsy*** group have the sequence PR, RT, RH, and IN.

In the years to follow the detection of the Ty1 element in yeast, further retrotransposons were characterized that were named Ty2–Ty5. The characteristics of these elements are discussed below. After completion of the yeast genome sequencing project (*cf.* Chapter 12), the complete list of the retrotransposons occurring in strain αS288C was established (Table 5.6).

5.1.5.2.2 Behavior of Ty Elements

Ty1 elements were found to be integrated into upstream regions of genes as well as into the coding regions of genes (Silverman and Fink, 1984). Moreover, in accord with their capability of transposition, they could be moved to new chromosomal loci into pre-existing Ty1 elements by a gene conversion mechanism (Roeder and Fink, 1980; Roeder *et al.*, 1980; Roeder and Fink, 1982) or be excised from a given chromosomal locus leaving behind only one of their LTR sequences, called *delta* elements. Such "solo" *deltas* were detected in many copies throughout the genome. Surprisingly, in the context of studying the organization of yeast tRNA genes, we provided evidence that Ty1 and *delta* sequences were often found associated with tRNA genes (Eigel and Feldmann, 1982; Baker *et al.*, 1982; Hauber *et al.*, 1988) and it has been confirmed that the 5′-flanking sequences constituted preferred integration sites for Ty transposition; in many cases, multiple integration and excision events were documented across the genome (Feldmann, 1988; Voytas and Boeke, 1993; Kim *et al.*, 1998; Hani and Feldmann, 1998).

Like retroviruses, the Ty elements transpose through an RNA intermediate and by reverse transcription (Boeke *et al.*, 1985). The retrovirus-like gene organization in Ty1 also became evident from its complete nucleotide sequence (Clare and Farabaugh, 1985; Hauber, Nelböck-Hochstetter, and Feldmann, 1985).

Transposition rates are low (one per 10^{-4} generations), and the number of elements is kept fairly constant by balancing transposition and excision events. This is manifest from the presence of 268 solo LTRs or other remnants that are footprints of previous transposition events. Due to the vagabond lifestyle of the retrotransposons, yeast strains differ with respect to the sometimes rather complex "patterns" formed by these elements resulting from multiple integrations and excisions. However, comparison of different yeast strains (e.g., Hauber *et al.*, 1986; Lochmüller *et al.*, 1989) and experimental data (Ji *et al.*, 1993) revealed that spontaneous transposition events do not appear to occur randomly along the length of individual chromosomes. Since these regions do not contain any special DNA sequences, the region-specific integration of the Ty elements may be due to specific interactions of the Ty integrase(s) with the transcriptional

complexes formed over the intragenic promoter elements of the tRNA genes or triggered by positioned nucleosomes in the 5′-flanking regions of the tRNA genes (e.g., Kirchner *et al.*, 1995). In any case, the Ty integration machinery can detect regions of the genome that may represent "safe havens" for insertion, thus guaranteeing both survival of the host and the retroelement.

5.1.5.2.3 **Expression of Ty Elements** Expression of the genes encoded within the Ty elements starts from a large transcript, which is produced by cellular RNA polymerase II; promoter elements regulating Ty transcription have been located to the 5′-LTR and to the internal part of the element, signals for transcriptional termination are close to the 3′-end of the internal region (*cf.* Figure 5.17).

First indications that Ty elements represent autonomous genetic entities that direct expression of endogenous genes was obtained from experiments in the Kingsmans' laboratory (Bowen *et al.*, 1984; Dobson *et al.*, 1984). Soon it was established that Ty1 followed a retrovirus-like strategy for the expression of a large fusion protein (Mellor *et al.*, 1985). Concomitantly, the second class of variant Ty elements, Ty2, was shown to obey a similar sequence organization and expression strategy as the Ty1-type elements (Fulton *et al.*, 1985).

Translation of the Ty mRNA initiates close to its 5′-end. Like many retroviruses, the elements Ty1–Ty4 employ translational frameshifting and rare tRNAs to regulate the expression of their gene products. In the Ty1/2 elements, two ORFs, TYA and TYB, comprise sequences encoding the retrovirus-like *gag* and *pol* proteins, respectively, whereby a translational frameshift (in a +1 mode) has to occur in the region overlapping TYA and TYB (Clare and Farabaugh, 1985; Wilson *et al.*, 1986), thus producing a *gag–pol* polyprotein. The minimal site for ribosomal frameshifting in Ty1/2 was determined to be a 7-nucleotide sequence (CUUAGGC) that induces tRNA slippage involving a minor tRNA species – a particular $tRNA^{Leu}$ that can read two different codons, CUU and UUA, so that "tRNA slippage" within the heptanucleotide can occur if translation is continued by $tRNA^{Gly}$ (Belcourt and Farabaugh, 1990). No tRNA slippage will occur if translation is continued by a $tRNA^{Arg}$ and run into a stop codon downstream of this site (Figure 5.18). This finding rendered an explanation at the molecular level as to why the *gag* versus *pol* protein precursors were produced in a ratio of 20 : 1 – translation of TYA was stopped at a usual stop codon in this minimal site, while read-through by frameshifting was limited by the availability of the rare tRNA. Similarly, +1 frameshifting has to occur in Ty4 expression, as the same sequence in the overlap between TY4A and TY4B was found as in Ty1/2 (Stucka *et al.*, 1989). In Ty3, the heptanucleotide overlap reads AUUAGUA.

Subsequently, the precursor polypeptides are processed by the endogenous protease (Figure 5.17), which like the endogenous proteases in other retrotransposons and retroviruses employs an aspartic residue in its catalytic center. Finally, the various subunits are assembled into VLPs, accommodating RT, IN, and PR, and Ty RNA, which in retrotranscription is used as a template to generate double-stranded Ty DNA that can be integrated at new sites within the genome (see Section 5.1.5.2.2).

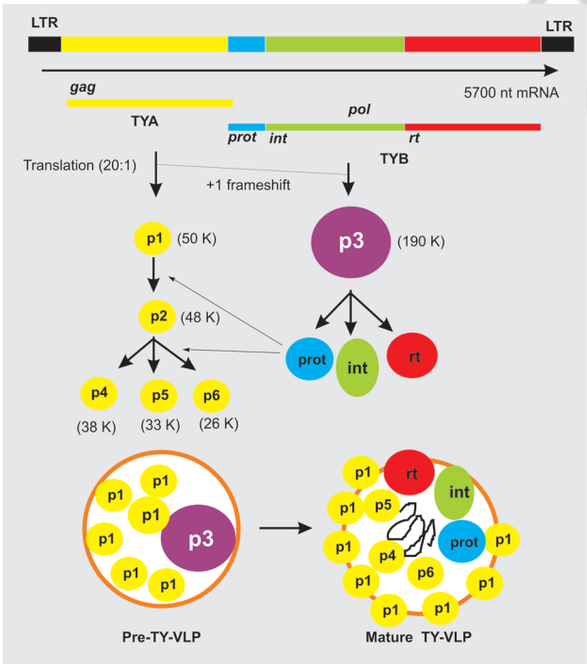


Fig. 5.17 Ty1 structure and expression strategy. LTR, long terminal repeat; *gag*, group-specific antigen (capsid); *prot*, protease; *int*, integrase; *rt*, reversed transriptase.

Codon usage in Ty ORFs. Codon usage in Ty expression is similar to that of the average of all yeast proteins, which means that the Ty elements fall into the category of intermediary expressed genes. However, for some codons there are substantial deviations towards the average codon preference in the 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) that the average (G + C) content of particular portions of the TYB proteins considerably differs from the average (G + C) content of host genes. As in highly expressed yeast genes, there is a tendency in one or the other element to avoid particular G/C-rich codons with a 3′ C or G.

More Ty elements. Although Ty1 and *delta* sequences were localized in region-specific distances upstream of tRNA genes, a novel repetitive element of about 330 bp, *sigma*, was found at the same positions in the 5′-flanking regions of two dissimilar yeast tRNA genes by Sandmeyer and Olson (1982). In the years to follow, it became evident that these insertions had occurred consistently in a 16–19 bp distance upstream of several tRNA genes (Brodeur, Sandmeyer, and Olson, 1983) and that the *sigma* elements constituted the LTRs of a novel class of yeast transposons – the **Ty3 elements** (Hansen, Chalker, and Sandmeyer, 1988).

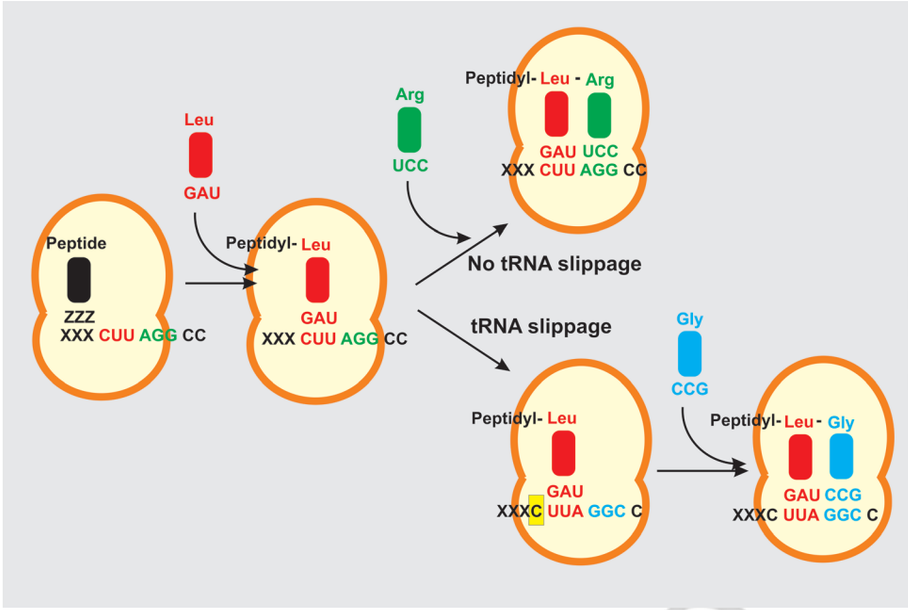


Fig. 5.18 Frameshifting during translation of Ty1 RNA.

The detailed characterization of Ty3 revealed (Hansen, Chalker, and Sandmeyer, 1988) that this element also transposes via VLPs as transposition-competent particles and exhibits translational frameshifting in a +1 mode. However, in comparison to Ty1 and Ty2, which belong to the *copia* family of retroelements initially found in *Drosophila*, Ty3 parallels the *gypsy* family of retroelements: this element carries a third ORF at its 3′-end and the sequence of the functional domains within the TY3 *pol* segment deviates from those in Ty1/2 (PR–IN–RT) in that the order in Ty3 is changed to PR–RT–IN. For the first time, experiments demonstrated that the association of Ty3 with tRNA genes reflects an interaction between the element and the RNA polymerase III transcription complex, which appears to direct integration (Chalker and Sandmeyer, 1992). More explicitly, in this case interactions between the N-terminal domain and TFIIIC subunit Tfc1p have been documented *in vitro*. and are consistent with *in vivo* results (Aye *et al.*, 2001). Recently, the function of the Ty3 nucleocapsid has been solved (Sandmeyer and Clemens, 2010).

A further retroelement, **Ty4**, was identified as a new class of yeast elements occurring in low copy number, belonging to the class of *copia* elements, and possessing a gene organization and expression strategy similar to Ty1/2 (Stucka, Lochmuller, and Feldmann, 1989); Ty4 also integrates into tRNA upstream regions.

The last retrotransposon found in yeast, **Ty5**, reveals a number of features deviant from those of the other Ty elements: its preferred target sites were identified to be silent chromatin regions, such as origins of replication at the telomeres and silent mating-type loci (Zou, Mitchell, and Stillman, 1997). Targeting was found to be mediated by interactions between Ty5 integrase and silencing proteins (Xie *et al.*, 2001), and it was argued that recognition of specific chromatin domains may be a general mechanism by which retrotransposons and retroviruses determine integration sites. The single ORFs present in the eight Ty5 elements from *S. cerevisiae* bear multiple stop codons, so that these Ty5 elements no longer appear capable of transposition. However, some strains of *Saccharomyces paradoxus* have numerous Ty5 insertions, suggesting that transposition is occurring in this species (Irwin and Voytas, 2001). Recently, the group of Voytas has used Ty5 in *S. cerevisiae* to show how specificity of integration is controlled in this particular case (Zhu, Bilgin, and Snyder, 2003). A commentary by Sandmeyer (2003) discusses these findings in the light of the various types of host–virus interactions that operate in different systems.

5.1.5.3 Ty Replication

Ty elements multiply like retroviruses. Like retroviruses, the yeast Ty elements replicate through RNA intermediates, and alternate their genetic material between RNA and DNA (Figure 5.19). **Reverse transcription** converts the genomic RNA into double-stranded DNA. Synthesis of each strand of retrotransposon DNA begins with the synthesis of short DNA products called minus-strand and plus-strand strong-stop DNA. Minus-strand strong-stop DNA synthesis is initiated from the 3′-hydroxyl group of a primer tRNA annealed at a primer-binding site (PBS) located just downstream of the so-called R-U5 sequence of the genomic RNA. Plus-strand strong-stop DNA synthesis begins at an RNase H-resistant oligoribonucleotide spanning a polypurine tract (PPT), located just upstream of the 3′-U3-R sequence of the RNA. Minus-strand and plus-strand strong-stop DNAs are elongated after they have been shifted to an acceptor region at the other end of the template (strand transfer). As a consequence, the unique 3′-U3 RNA sequence is duplicated at the

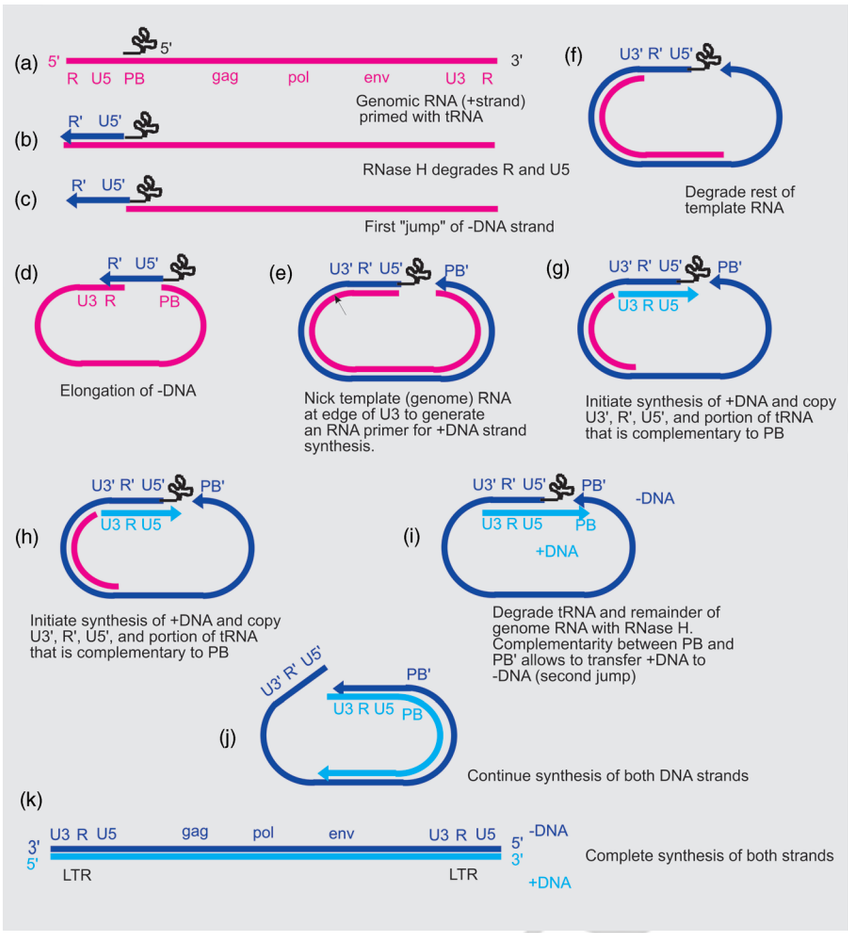


Fig. 5.19 Steps in Ty replication.

5′-end of the preintegrative DNA to form the upstream LTR and the unique 5′-U5 RNA sequence is duplicated at the 3′-end of the DNA to form the downstream LTR. The final product of reverse transcription is a two-LTR linear double-stranded DNA molecule that is longer than the genomic RNA. For most retroviruses the replicated linear extrachromosomal DNA has 2 bp at each end that are not present at the end of integrated proviral DNA. The first step in retroviral integration is the cleavage of 2 nucleotides from the 3′-end of preintegrative DNA. The replicated DNA integrated into the genomic DNA of the host cell can be transcribed to produce new molecules of genomic RNA. The basis for retroviral replication had already been laid by Varmus and colleagues (Majors *et al.*, 1981; Varmus, 1982), and the above scheme can be found in Wilhelm *et al.* (1997). It may be useful for the reader to learn more on this issue from some articles that have appeared since (Karst, Rutz, and Menees, 2000; Irwin *et al.*, 2005; Lemoine *et al.*, 2005; Resnick, 2005).

5.1.5.4 Interactions between Ty Elements and their Host

Where to insert? One recurrent theme after regulation of transposition remained **target site selection** (Curcio, Sanders, and Garfinkel, 1988). Retrotransposon Ty1 faces a formidable cell barrier during transposition – the yeast nuclear membrane, which remains intact throughout the cell cycle. Therefore, it was an interesting finding that Ty1 integrase has a nuclear localization signal at its C-terminus, to substantiate the mechanism by which transposition intermediates are transported from the cytoplasm to the nucleus, where they are integrated into the genome (Kenna *et al.*, 1998).

Although target site selection is still not well understood for this general class of retroelements, it became clear over the years that Ty elements target their integration to very specific regions of their host genomes, as revealed by the analysis of genomic integration sites in the group of J. Boeke. Targets containing genes transcribed by RNA polymerase III were found several 100-fold more active as integration targets for Ty1 than "cold" sequences lacking such genes. High-frequency targeting depended on polymerase III transcription and integration was found "region specific," occurring exclusively upstream of the transcription start sites of these genes in a window of around 700 bp (Devine and Boeke, 1996). The pattern of insertion **upstream of tDNA** was non-random and not distributed equally throughout the genome,

but periodic, with peaks separated by round 80 bp (Bachman, Eby, and Boeke, 2004). It has been demonstrated that ATP-dependent chromatin remodeling by Isw2p upstream of tRNA genes leads to changes in chromatin structure and Ty1 integration site selection, and that Bdp1p, a component of the RNA polymerase III transcription factor TFIIIB, is required for targeting Isw2 complex to tRNA genes (Bachman *et al.*, 2005).

Interestingly, it was recently demonstrated that the Ty element itself has a closed (nuclease-insensitive) **chromatin configuration** that is also imposed on the flanking DNA sequences. The compact chromatin structure is determined by sequences at the N-terminus of the Ty. Increased binding of the Rap1 protein to the hotspot restores both open chromatin conformation and DSB formation. The chromatin configuration of Ty elements precludes initiation of recombination, thus preventing potentially lethal exchanges between repeated sequences (Ben-Aroya *et al.*, 2004).

Conflict of interests or benefit? The second question, whether there is a transcriptional interference between Ty insertions and tRNA genes, had been answered positively by our experiments using a unique "artificial tRNA gene" (*SYN2*) that was tagged by an intron-like sequence that could not be spliced out from its long precursor, but otherwise behaved like resident tRNA genes (Krieg *et al.*, 1991). This gene combined with various Ty constructs and integrated as a single copy each into the yeast genome was used to monitor the transcriptional interference between Ty (and segments thereof) and a flanking tRNA gene as well as the chromatin conformation of the stable transcription complex and its flanking regions (Feldmann, 1988; Krieg *et al.*, 1991). A modest stimulatory effect was observed (like in the majority of regulatory systems in yeast) of Ty or LTR insertions upstream of a tRNA gene on its expression *in vivo*. Transcriptional interference between Ty1 insertions and two polymerase III-transcribed genes was later also shown in the cases of tagged *SNR6* and *SUP2* (Bolton and Boeke, 2003); *vice versa*, RNA analysis indicated a modest tRNA position effect on Ty1 transcription at native chromosomal loci. Furthermore, this study revealed that tRNA genes exert a modest inhibitory effect on adjacent polymerase II promoters, a result that was confirmed in other experiments (Wang *et al.*, 2005b).

The problem of correlating tRNA gene expression with chromatin structure was more complex. The data mentioned above (Feldmann, 1988; Krieg *et al.*, 1991 and references cited therein) supported the following model. (i) tRNA genes counteract the formation of a canonical chromatin structure over a window reaching from around 30 bp each upstream and downstream. In other words, actively transcribed tRNA genes have to be kept free of nucleosomes. (ii) The general pattern tRNA genes exhibit in DNase I digestion experiments is a triplet of hypersensitive sites resulting from protection of sequences at the A- and B-box elements and accessibility upstream and downstream from the structural gene and between A- and B-boxes, reflecting the binding of TFIIIC to the intragenic promoter and the tight binding of TFIIIB to the upstream transcription initiation site (around 30 bp in length; *cf.* Section 9.2.2). (iii) Accessibility of this site by TFIIIB is crucial for active tRNA gene transcription, so that this region has to be kept in a nucleosome-free configuration. (iv) In DNase I experiments, the adjacent hypersensitive site(s) indicating canonical nucleosome spacing are located around 170 and around 340 bp upstream from the initiation start site of actively transcribed tRNA genes. The first upstream nucleosome in these instances is found positioned in such a way as to form a boundary induced by the transcription complex. (v) A prerequisite for the induction of such a constellation is that the formation of the transcriptional complex outweighs the formation of nucleosomes – a situation that prevailed in competition experiments. (vi) Whenever the sequences upstream of a tRNA gene are "favorable" to assist this positioning effect, transcription is enabled at a normal or even slightly elevated level. In "unfavorable" cases, however, nucleosomes can be formed over these sequences, thus exerting a constraint for transcriptional initiation. The highest transcriptional rates were always found in constructs, in which Ty elements, *delta* or *tau* sequences, had been placed into "native" distances upstream of a tRNA gene, rendering "favorable" constellations.

Very recently, the nucleosomal occupancy and expression rate of the SUP4 tRNA gene has been investigated. The authors (Mahapatra *et al.*, 2011) arrived at similar conclusions as to the location of one upstream and one downstream nucleosome as discussed above. A positioned nucleosome is found between –192 to –47 upstream of the tRNA structural part (commencing at position +1) and a more flexible nucleosome extending from +98 to +242 up to +220 to +365 downstream, leaving a nucleosome-free, polymerase III-transcribed region. Both these nucleosomes contain the H2A.Z variant histone. The two remodeling complexes FACT and RSC play different roles in transcription regulation. FACT appears to exert an inhibitory role, while RSC helps to keep the gene nucleosome-free and under stress conditions activates transcription by shifting the nucleosome abutting the terminator under normal conditions more downstream.

5.2 Yeast tRNAs, Genes, and Processing

5.2.1 Yeast tRNAs

5.2.1.1 Yeast Led the Way to tRNA Structure

The first small RNA molecules under investigation. The **adaptor hypothesis**, formulated by Francis Crick in 1957 (Crick, 1957) in connection with his thoughts on the genetic code (Crick, 1966), proposed that during protein synthesis the single amino acids concatenated to a peptide chain on "microsomal particles" are carried by specific adaptor molecules. On the one hand, these adaptors can form a stable

bond with specific amino acids and, on the other hand, these are capable of delivering these amino acids to the growing peptide chain by reading the information from a (microsomal) RNA template according to the same base-pairing rules as found in DNA. In the same year, Hoagland, Zamecnik, and Stephenson (1957) reported the discovery of what then was collectively called sRNA (soluble ribonucleic acid) and what we now call tRNA (transfer ribonucleic acid), after having shown that the activation of amino acids for the RNA-dependent synthesis of polypeptides involves the cleavage of ATP to AMP and pyrophosphate, with the intermediate formation of aminoacyl-AMP compounds.

Using sRNA from *Escherichia coli* or liver cell extracts, Fritz Lipmann's group could show that the amino acids were, in fact, chemically bound via a highly reactive aminoacyl ester bond to the 3′-terminus of these molecules (Zachau, Acs, and Lipmann, 1958). The **aminoacylation test** (Hoagland, Zamecnik, and Stephenson, 1957) became an excellent means to demonstrate that particular amino acids were bound to specific sRNA components. At this time, however, it was nearly impossible to fractionate sRNA from cell extracts into single species to yield individual tRNAs.

The first success of **isolating tRNAs** free from other cellular components seems more than fortuitous. Zamecnik and his colleagues (Monier, Stephenson, and Zamecnik, 1960) had noticed that direct extraction of yeast with aqueous phenol could be used for this purpose because under the conditions employed, little of the high-molecular-weight material was released from the cells. This method could be applied for large-scale preparation of total tRNA from yeast. One early application based on this approach was the large-scale preparation of yeast aminoacyl-tRNA, which was then used in our experiments to establish that the amino acid was preferentially linked to the 3′-OH moiety of the terminal adenosine residue (Feldmann and Zachau, 1964). More importantly, large-scale preparation of tRNA formed the basis to obtain purified amino acid-specific tRNAs for biochemical analysis. Experiments along this line started in the late 1950s. One has to recollect, however, that working out appropriate fractionation procedures and applying them for mass production was a hard task. First, tons of yeast slurry had to be subjected to phenol extraction, and raw tRNA had to be precipitated from the aqueous phase with ethanol (or acetone) and further purified by column chromatography on DEAE cellulose. This scale of mass preparation of starting material was later by far surpassed when Kornberg and his colleagues "manufactured" 10 000 l of yeast cultures to obtain sufficient quantities of RNA polymerase II transcription complexes and transcription factors for X-ray studies (Darst *et al.*, 1991; Bushnell and Kornberg, 2003).

The **isolation of amino acid-specific tRNA** was the most tedious task: fractionation of total tRNA by means of consecutive countercurrent distributions in various systems, column chromatography, and so on, whereby each single fraction had to be measured for amino acid acceptor activity. Of course, the subsequent analytical procedures to be applied (partial and complete digestion with more or less specific nucleases, fractionation of the fragments or components, determination of the nature of the single constituents) extended over several years. However, in the end, Holley and his coworkers in the United States succeeded in determining the first sequence of a tRNA – the alanine-specific tRNA from yeast (Holley *et al.*, 1965) (Figure 5.20a). Soon after, Hans Zachau's group from the new Institute of Genetics in Cologne reported the sequences and the structures of the modified nucleotides from two yeast serine-specific tRNAs (Zachau, Dütting, and Feldmann, 1966; Zachau *et al.*, 1966) (Figure 5.20b).

The next tRNA sequences to be elucidated were those of yeast $tRNA^{Tyr}$ (Madison, Everett, and Kung, 1966) and $tRNA^{Phe}$ (Raj Bhandary *et al.*, 1966). In the years to follow, the sequences of a great variety of amino acid-specific tRNAs from yeast and some of their isoacceptors were determined in several laboratories. The groups of Guy Dirheimer at the Institute of Molecular and Cellular Biology and the Faculty

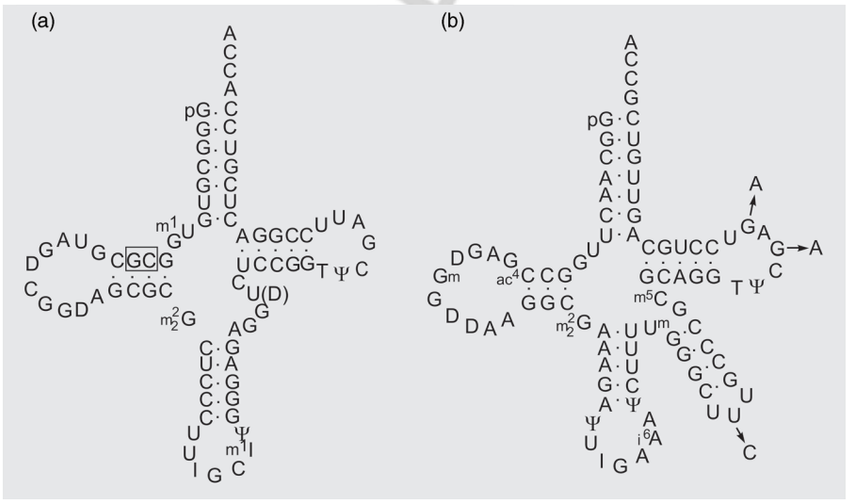


Fig. 5.20 (a) Clover-leaf structure of the alanine-specific tRNA (Holley *et al.*, 1965). The boxed G–C has to be removed according to Penswick, Martin, and Dirheimer (1975). (b) Cloverleaf structure of the serine-specific tRNAs (Zachau, Dütting, and Feldmann, 1966a; Zachau *et al.*, 1966b). Substitutions in Ser tRNA I versus Ser tRNA II are marked by arrows.

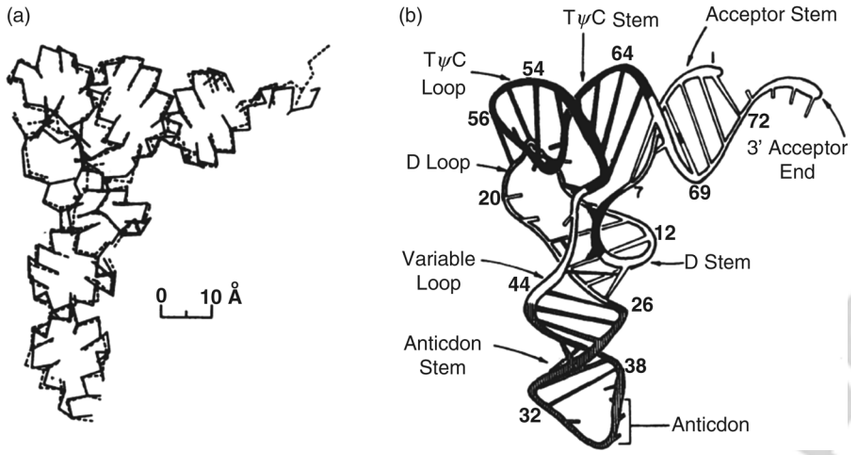


Fig. 5.21 Three-dimensional structure of yeast phenylalanine tRNA. (a) The solid line connecting the group coordinates represents the conformation of the molecule in the orthorhombic unit cell, while the dashed line shows its conformation in the monoclinic unit cell. (b) Secondary and tertiary hydrogen bonds between bases are shown with different shading. The numbers refer to the residues in the polynucleotide chain. (Reproduced from Quigley *et al.*, 1975, with permission from Oxford University Press.)

of Pharmacy in Strasbourg contributed many yeast nuclear-encoded tRNA structures – a subject they pursued from 1968 until the end of the 1970s (Dirheimer, 2005); sequencing in Strasbourg was continued to characterize yeast mitochondrial tRNAs, mainly in the decade to follow (Dirheimer, 2005; Dirheimer and Martin, 1990).

A plethora of **modified nucleosides** occurring in yeast nuclear and mitochondrial tRNAs were identified and characterized by many workers as well. Characterization of the modified nucleosides was facilitated by the mass preparation of specific tRNAs and more than 50 could be detected in yeast (Sprinzl *et al.*, 1996). In most cases, it was also possible to delineate the enzymatic pathways by which the "odd" compounds are derived from their parent nucleotides. Many laboratories were involved in this research and this field is still under investigation as a few citations may illustrate (Bjork *et al.*, 1987; Dihanich *et al.*, 1987; Ellis, Hopper, and Martin, 1987; Edqvist, Grosjean, and Straby, 1992; Grosjean *et al.*, 1996; Gerber and Keller, 1999; Anderson, Phan, and Hinnebusch, 2000; Gerber and Keller, 2001; Schaub and Keller, 2002; Bjork *et al.*, 2001; Xing *et al.*, 2004; Martin, Doublié, and Keller, 2008). Finally, RNA modifications could be discovered by using microarrays (Hiley *et al.*, 2005).

A rational extension of the work on yeast tRNA structure was to solve the problem of how aminoacyl-tRNA synthetases specifically interact with their cognate partners in the aminoacylation reaction (Lengyel and Soll, 1969). This afforded the purification of specific enzymes and allowed the set-up of appropriate methods for the isolation of tRNA/synthetase complexes to carry out X-ray studies and fast kinetic techniques to determine reaction parameters. Important contributions came from a most successful cooperation between the groups of Ebel, Giegé, Moras, Grosjean, and others (Ebel *et al.*, 1973; Moras *et al.*, 1983; Giegé *et al.*, 1990). This is not to deny that other organisms, mainly *E. coli*, have contributed a lot to this field. In all, these investigations led to the important notion that each tRNA/synthetase system has to follow intrinsic rules for recognition given by the conformational features of the partners. **tRNA "identity"** was also recognized as the superior criterion for the interactions between (pre)tRNAs and the nucleotide-modifying enzymes.

Remarkably, the first **three-dimensional structure** of a tRNA molecule was also derived from a yeast tRNA – the phenylalanine-specific tRNA – by Alex Rich and coworkers in 1974–1975 (Rich and Kim, 1978). Figure 5.21 shows the comparison of the conformation of yeast phenylalanine tRNA in two crystal forms. The two molecules from the orthorhombic and monoclinic unit cells have been fitted by a least-squares procedure. Three group coordinates are plotted: the position of the phosphorus atom, the centroid of the five atoms in the furanose ring of ribose, and the centroid of the six atoms that make up the six-membered ring in either pyrimidines or purines.

The shape of its three-dimensional structure caused Francis Crick to compare it to the folded structure of protein saying that "tRNA mimics a protein." More three-dimensional structures of tRNAs were disclosed later (e.g., by the approaches of Sigler and his collaborators (Basavappa and Sigler, 1991)).

5.2.1.2 Yeast tRNA Precursors and Processing

Mature tRNAs are generated from precursors. During the 1970s, attention was drawn to the cellular processes as to how tRNAs (the prototypes of simple and stable RNAs) are transcribed from the corresponding genes and how the resulting precursors are processed to finally result in their mature form. Necessary steps in **tRNA biosynthesis** included modifications of particular nucleotides and, in eukaryotes, the enzymatic addition of the universal 3′-CCA end not contained in the gene sequences to the "core" tRNA (Deutscher, 1975; Ghosh and Deutscher, 1980).

From the pioneering work of Darnell on RNA precursors, it was already manifest that eukaryotic tRNA precursors must contain additional sequences at their 5′- and 3′-ends that have to be removed during maturation (Bernhardt and Darnell, 1969). Using extremely short pulses of ^{32}P during yeast growth and subsequent fractionation of the precursors

by polyacrylamide gel electrophoresis we noted that some specific precursors were considerably longer than the matured molecules and that processing might occur in consecutive steps (Blatt and Feldmann, 1973). Fractionation and characterization of a total population of specific tRNAs and **tRNA precursors** was later refined by a two-dimensional gel electrophoresis approach (Fradin, Gruhl, and Feldmann, 1975). In contrast to sea urchin tRNA genes found to be arranged in clusters of tandem repeats (Clarkson, Birnstiel, and Purdom, 1973) or most of the *E. coli* tRNA genes being arranged in a polycistronic mode (Altman, 1975), the yeast tRNA genes were found to occur as single transcriptional units scattered throughout the genome (Feldmann, 1977). The only exception to this rule later turned out to be a yeast $tRNA^{Asp–Arg}$ pair (Schmidt *et al.*, 1980). Many precursors to specific tRNAs were identified by Hopper and coworkers (Hopper and Kurjan, 1981).

A big surprise came from the sequence analysis of a yeast $tRNA^{Tyr}$ gene and the analysis of its transcript by Goodman, Olson, and Hall (1977): the $tRNA^{Tyr}$ precursor revealed the presence of a 14-nucleotide **intervening sequence** located 3′ to the anticodon, being removed during the maturation process. Thus, the cloning and analysis of the first tRNA genes from yeast indicated that split genes occurred not only in adenovirus (Berget, Moore, and Sharp, 1977; Gelinas and Roberts, 1977), but, as found soon after, also in protein-encoding genes in mammalian cells (Breathnach, Mandel, and Chambon, 1977; Jeffreys and Flavell, 1977; Tilghman *et al.*, 1978). The first intron in a yeast mitochondrial gene (the large rRNA coding gene) was identified in 1978 (Bos, Heyting, and Borst, 1978), although its presence was so obvious from the differences between ω^+ and ω^- strains as observed by Dujon in 1974 (see Section 11.1).

In the early 1980s, more yeast **tRNA genes** containing introns were characterized and studies on their maturation begun (Abelson, 1980; Ogden *et al.*, 1980; Valenzuela *et al.*, 1980). Finally, it became evident that some 25% of the yeast tRNA genes carry introns of variable length, but always at the same position – next to the 3′-side of the anticodon. It was mainly John Abelson and his collaborators who became interested in yeast tRNA splicing, and after many years succeeded in unraveling the details of this maturation step and in characterizing the endonucleases involved in this process (Abelson, Trotta, and Li, 1998). An interesting observation was that in some cases modifications of particular nucleotides depended on the presence of the intron sequences. However, the functional significance of these introns (occurring generally in eukaryotic tRNA genes) largely remains a mystery.

5.2.2 Current Status of Yeast tRNA Research

5.2.2.1 Yeast tRNAs and their Genes

Getting the complete repertoire. The determination of the entire sequence of the yeast genome (*cf.* Chapter 12) revealed the occurrence of **274 tRNA genes**, which are scattered throughout the genome (Table 5.7). According to their amino acid acceptor activities, the tRNAs (and their genes) can be grouped into 42 families. Families for the major tRNA species comprise up to 12 copies, while minor tRNA species are restricted to one to three gene copies. The peculiarity of yeast tRNA genes to be associated with Ty elements has been discussed in Section 5.1.5 above. Recent experiments have indicated that tRNA genes are preferentially localized to the nucleolus, as has also been found for other eukaryotes (Thompson *et al.*, 2003; Haeusler *et al.*, 2008). Even in yeast, however, not every individual tRNA gene is likely to be localized to the nucleolus at all times. Mapping of a limited number of tRNA gene loci revealed 60–70% nucleolar association if the tRNA gene was actively expressed versus 5–15% if transcriptionally inactivated, but it is possible that positioning is quite variable among loci.

Transcription of the tRNA genes can be described in sufficient detail; in Section 9.2.2 we discuss how RNA polymerase III and several transcription factors bind to the **internal A and B promoter boxes**, and to the upstream noncoding sequences. An extremely surprising finding in the analysis of the yeast tRNA genes was that practically no sequences in the flanking regions outside the structural part of the genes have been conserved; this finding holds true even for the redundant copies of a particular tRNA gene. No plausible model has been suggested that would explain how multiple copies of a unique tRNA gene are generated and distributed to many locations across the genome. A further problem that has never been solved satisfactorily concerns the expression level of individual tRNA genes. Although it has been possible to measure *relative* levels of **tRNA gene expression** depending on the flanking regions, there are no means of comparing the *absolute* expression levels among different copies of the same tRNA gene. A vague impression that differences in expression must exist came from experiments, in which individual copies of the five initiator $tRNA_i^{Met}$ genes were deleted in different combinations and the survival of the cells was measured. It turned out that some combinatory deletions were more hazardous than others. Using microarrays it is now possible to determine the level of total tRNA aminoacylation under varying conditions, but this technique is not suitable to measure charging of individual tRNAs (Zaborske and Pan, 2010).

5.2.2.2 tRNA Processing and Maturation

Maturation is a complex process. The processing pathways for pre-tRNAs have unique constraints in eukaryotic nuclei. The many tRNAs require a huge diversity of different types of processing to refine their structure and identity (Figure 5.22). One important component of the early tRNA processing machinery is RNase P (review: Walker and Engelke, 2006) – an endonuclease complex that catalyzes the cleavage of the 5′-leader sequence from pre-tRNA transcripts. RNase P is thought to be a primordial enzyme dating

Table 5.7 Yeast tRNA families and their genes.

tRNA species[a)]	Number of functional genes[b)]	tDNA anticodon	tRNA anticodon[c)]	Probable codon preferences[d)]	Remarks on variant tRNA or tDNA sequences[e)]
Ala1	11	AGC	IGC	GCU; GCC	
Ala2	5	TGC	*UGC	GCA; GCG?	
Arg1	1	CCT	CCU	AGG	
Arg2	6	ACG	ICG	CGU; CGC; CGA?	
Arg3	11	TCT	^{mcm5}UCU	AGA	four dimeric (Arg–Asp) genes; no gene for a variant tRNA(Arg3) in αS288C
Arg4	1	CCG	*CCG	CGG	
Asn	10	GTT	GUU	AAU; AAC	
Asp	15	GTC	GUC	GAU; GAC	four dimeric (Arg–Asp) genes
Cys	4	GCA	GCA	UGU; UGC	
Gln1	7	TTG	*UUG	CAA	
Gln1a	2	TTG	*UUG	CAA	these variants to tRNA^{Gln1} have three alternate bases in the AC stem
Gln2	1	CTG	*CUG	CAG	essential gene, closely related to Gln1
Glu3	14	TTC	^{mcm5s2}UUC	GAA	
Glu4	2	CTC	*CUC	GAG	
Gly1	16	GCC	GCC	GGU; GGC	
Gly2	3	TCC	NCC	GGA	
Gly3	2	CCC	*CCC	GGG	
His2	7	GTG	GUG	CAU; CAC	
Ile1	2 (i)	TAT	*UAU	AUA	probably one variant gene
Ile2	13	AAT	IAU	AUU; AUC	
Leu1	3 (i)	TAG	UAG	CUA; CUG? CUU; CUC UUA; UUG	variant intron sequences in one gene copy
Leu3	10(i)	CAA	^{m5}CAA	UUG	variant introns in five gene copies
Leu4	7	TAA	NAA	UUA; UUG	
Leu5	1	GAG	*GAG	CUU; (CUC?)	
Lys1	14	CTT	CUU	AAG	probably two variant genes
Lys2	7 (i)	TTT	^{cmnm5}UmUU	AAA	variant intron in one gene copy
Meti	5	CAT	CAU	AUG	
Met3	5	CAT	CAU	AUG	variant tRNA^{Met3} observed in other strain
Phe	8 (i)	GAA	G^{m}AA	UUU; UUC	variant intron sequences
Phe1a	2 (i)	GAA	G^{m}AA	UUU; UUC	tRNA^{Phe1a} has alternate bases in acceptor stem; genes have variant introns
Pro1	10 (i)	TGG	?UGG	CCA; CCG?	one variant gene; variant introns in five gene copies
Pro2	2	AGG	probably IGG	CCU; CCC	
Ser2	11	AGA	IGA	UCU; UCC	
Ser3	4 (i)	GCT	*GCU	AGU; AGC	two genes with variant AC loop (*cf.* text)
Ser4	3	TGA	?UGA	UCA	
Ser5	1 (i)	CGA	*CGA	UCG	
Thr1a	11	AGT	IGU	ACU; ACC	
Thr2	1	CGT	*CGU	ACG	
Thr3	4	TGT	*UGU	ACA	probably one variant gene
Trp	6 (i)	CCA	C^{m}CA	UGG	
Tyr	8 (i)	GTA	GψA	UAU; UAC	variant introns in three gene copies
Val1a	13	AAC	IAC	GUU; GUC	
Val1b	1	AAC	IAC	GUU; GUC	variant tRNA^{Val1} gene
Val2a	2	TAC	^{ncm5}UAC	GUA	
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 Czerwoniec *et al.* (2009). An asterisk indicates that this base has been deduced from the DNA sequence, but that further information on whether this 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. Largely, we follow the proposed conventions. Question marks refer to the codons discussed in the text.
e) Variants giving rise to suppressor tRNAs are not listed here.

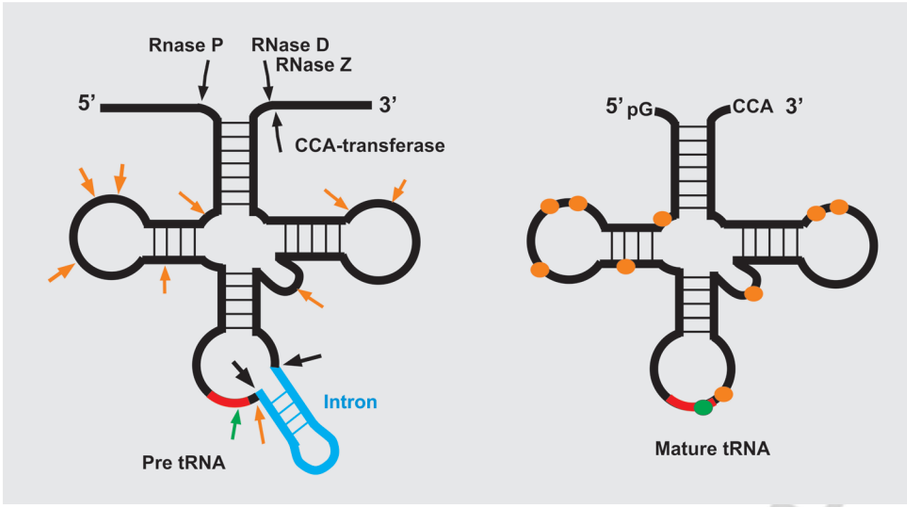


Fig. 5.22 Scheme for processing of tRNA precursors.

back to the "RNA world," as it exists as a RNP complex in almost all organisms with a large, catalytic RNA and variable amounts of auxiliary protein. In budding yeast, both RNase P and early pre-tRNA transcripts, which contain 5′- and 3′-termini and introns, primarily localize to the nucleolus (Bertrand *et al.*, 1998). In addition, there is some evidence that RNase P might associate directly to the polymerase III transcription apparatus through interactions with transcription factor TFIIIB, which binds upstream of tRNA genes. Another early processing function common to all pre-tRNA transcripts – cleavage of the 3′-trailing sequences – usually occurs quickly after RNase P cleavage; one such enzyme is RNase Z (Dubrovsky *et al.*, 2004).

In yeast, genes encoding 11 different nuclear tRNAs contain introns: the species designated $tRNA^{Ile1}$, $tRNA^{Leu1}$, $tRNA^{Leu3}$, $tRNA^{Lys2}$, $tRNA^{Phe1}$, $tRNA^{Phe1a}$, $tRNA^{Pro1}$, $tRNA^{Ser3}$, $tRNA^{Ser5}$, $tRNA^{Trp}$, and $tRNA^{Tyr}$. The introns range from 14 to 60 nucleotides in length and interrupt the anticodon loop next to the 3′ position of the anticodon triplet. No conservation of sequence has been observed at the splice junctions; only the 3′-splice site is located in a bulged loop. No further "rules" can be established. Moreover, the sequence of an intron in a given tRNA species can vary and there is no explanation why in some cases only particular members in a tRNA gene family possess introns (e.g., see $tRNA^{Leu}$ or $tRNA^{Ser}$).

Excision of the introns from eukaryotic tRNA precursors is an enzyme-catalyzed process, which has been completely solved taking yeast as an example. Pre-tRNA splicing occurs in three consecutive steps. (i) The intron is removed by a heterotetrameric endonuclease (Sen2p, Sen15p, Sen34p, and Sen54p) producing a 5′-"half" molecule with a 2′–3′ cyclic phosphate residue and a 3′-"half" molecule with a 5′-OH group (Trotta *et al.*, 1997). It is thought that the endonuclease complex contains two active sites, each responsible for one of the cleavages. (ii) The second step is a ligation reaction carried out by the multifunctional tRNA ligase, Trl1p, which first adds a phosphate residue to the 3′-half, and then joins the 5′- and 3′-halves in a phosphoester bond (Phizicky *et al.*, 1992). (iii) In the final step, catalyzed by tRNA 2′-phosphotransferase (Tpt1p), the excess 2′-phosphate from the splice junction is removed and used to generate an usual byproduct, adenosine diphosphate-ribose 1″,2″-cyclic phosphate (Culver *et al.*, 1993). The tRNA splicing endonuclease subunits are conserved from Archaea to man; the crystal structure of an Archaea enzyme was published in 2009. Although tRNA splicing affects precursor molecules that would be directly available in the nucleus, in yeast neither the endonuclease nor the two consecutive enzymes are found in the nucleus, but surprisingly the endonuclease is found located on the cytoplasmic surface of mitochondria. A survey of tRNA splicing in different eukaryotes has revealed that the location of splicing differs, while the biochemistry of pre-tRNA processing is conserved.

Both nuclear and mitochondrial yeast tRNAs are highly modified from their precursors before they are released as mature tRNAs. More than 100 different **chemical modifications** and most of the corresponding enzymes have been described (overview: Czerwoniec *et al.*, 2009). These distinct modifications include numerous methylations of all four bases or the 2′-OH group of the ribose moieties, isomerization of uracil into pseudouracil, conversion of uracil into dihydrouracil (preferably in the "dihydro-U loop"), and *N*-acetylation of cytidine, deamination of adenine to inosine; more elaborate modifications are used for nucleotides within the anticodon (e.g., methoxycarbonylmethylation or thiolation of uracil) or adjacent to the anticodon at its 3′-side (e.g., isopentenylation of adenosine, formation of wybutosine or queuosine from guanosine). Although the functions of the modified nucleotides are not understood in detail, they appear to be necessary for maintaining tRNA stability and tRNA three-dimensional

structure; only modifications of anticodon bases are known to participate in decoding (see below).

Meanwhile, it has been established that the **tRNA modifications** occur in a preferred order. Some of the modifying enzymes employ only pre-tRNAs, such as intron-containing precursors, as substrates, while others can handle only spliced molecules. Further, some of the modifying enzymes are localized in the nucleus or in subnuclear locations, whereas others are solely bound to the cytoplasm. Presently, even in the yeast system, it seems nearly impossible to obtain information on the location of the modifying activities (an overview on modifying enzymes is given in Section 3.6.4). This notion invokes the view that the tRNAs have to travel through different cellular compartments to acquire the necessary modifications. Finally, it remains extremely difficult to imagine how this whole process becomes regulated (Phizicky and Alfonso, 2010).

5.2.2.3 Participation of tRNAs in an Interaction Network

There are more than 20 functional classes of tRNAs mediating protein synthesis. Apart from the canonical elongators, initiator tRNAs initiate translation, while noncanonical elongators mediate expansion of the genetic code to the cotranslationally inserted amino acids 21 and 22, selenocysteine and pyrrolysine. Within broad functional classes, isoacceptor tRNAs are targeted to interact with the various specific enzymes for covalent modification (see above) or mediate programmed translational events at the ribosome. There are still many other interactions – particularly in organisms other than yeast – carried out by specific tRNAs.

These diverse functions among tRNAs are achieved through specificity in RNA–protein interactions. The **functional "identity"** of a tRNA is laid down in unique structural features that allow for specific interactions with proteins, RNPs, and RNA complexes that catalyze particular steps in biosynthesis, maturation, modification, quality control, charging with amino acids, and targeting reactions (Ardell, 2010). The tRNA–protein interacting network has a hierarchical structure. Some interactions (e.g., with biosynthetic nucleases, EF-Tu, or the ribosome) are shared by many tRNA classes, whereas others are restricted to a few or only single tRNA species, like the aminoacylation reactions. To accommodate conflicts arising in such a hierarchy, tRNA need not only employ identity determinants, but also identity antideterminants that discriminate between interactions and block those unwanted.

5.2.2.3.1 Aminoacylation of tRNAs

Protein biosynthesis has been studied in yeast as one of the first eukaryotic model organisms, next to *E. coli*. Many basal findings on the structure and function of tRNAs, tRNA synthetases, 80S ribosomes, and the initiation, elongation, and termination factors mediating translation have been identified in yeast and studied in great detail (see also Chapter 12). In fact, some of the principles of protein biosynthesis and the function of the adaptor molecules in this process have been elucidated with the help of yeast. In aminoacylation, the concept of tRNA identity was developed rather early.

We will not elaborate here on the details of ribosomal protein biosynthesis, but it may be useful to briefly enumerate the various **aminoacyl-tRNA synthetases** charging the tRNAs with their cognate amino acids (Table 5.8). Aminoacyl-tRNA synthetases possess precise substrate specificity and, despite their similarity in function, vary in size, primary sequence, and subunit composition (Delarue, 1995; Arnez and Moras, 1997). Individual members of the aminoacyl-tRNA synthetase family can be categorized in one of two classes, depending on amino acid specificity (Eriani *et al.*, 1990). **Class I enzymes** (those specific for Glu, Gln, Arg, Cys, Met, Val, Ile, Leu, Tyr, and Trp) typically contain two highly conserved sequence motifs, are monomeric or dimeric, and aminoacylate at the 2′-terminal hydroxyl of the cognate tRNA (which is then transformed into the 3′-terminal hydroxyl form). **Class II enzymes** (those specific for Gly, Ala, Pro, Ser, Thr, His, Asp, Asn, Lys, and Phe) typically contain three highly conserved sequence motifs, are dimeric or tetrameric, and aminoacylate at the 3′-terminal hydroxyl of the cognate tRNA. It is noteworthy that – in contrast to the tRNAs – mitochondria receive their own mitochondrial-specific set of aminoacyl-tRNA synthetases. In a few cases (Ala, Gly, His, and Val), the same genes encode both types of synthetases, whereby the presequences of the corresponding proteins may differ.

It is well known that tRNA synthetases are engaged in **multiaminoacyl-tRNAsynthetase (MARS)** complexes, held together by auxiliary nonenzymatic factors; it is even speculated that these complexes are an anchoring platform for synthetases to fulfill other tasks than just aminoacylation (Hausmann and Ibba, 2008). Recently, a complex consisting of three proteins, cytosolic methionine tRNA synthase (MRS), cytosolic glutamyl-tRNA synthase (ERS), and Arc1p, has been investigated in yeast (Frechin *et al.*, 2010). Arc1p is a basic protein with a predicted molecular mass of 42 kDa. The protein sequence can be divided into three domains, an N-terminal GST-like fold (N domain) that interacts with MRS and ERS, a central nonspecific RNA-binding domain (M), and a C-terminal specific tRNA-binding domain (C domain); the M and C domains together make up the TRBD (tRNA-binding domain). In the complex, the TRBD works in tandem with the anticodon-binding domain of the two synthetases, which is crucial for recognition and discrimination of the cognate tRNA. Additionally, a putative functional homolog of Arc1p, Cex1p, has been shown to be involved in tRNA channeling from the nuclear pore to the translational machinery. At the nuclear side, Utp8p functions in transferring tRNAs to the export receptors. In all, one combined task of these factors is to exert a general control on tRNA traffic.

In yeast, as well as in any other eukaryote, the gene encoding mitochondrial glutaminyl-tRNA synthetase is missing, suggesting that the synthesis of mitochondrial glutaminyl-$tRNA^{Gln}$ is achieved via the transamidation pathway. A further peculiarity of yeast mitochondria is that under fermentation conditions only basal levels of mitochondrial activity are

Table 5.8 Cytosolic and mitochondrial aminoacyl-tRNA ligases in yeast.

Amino acid	Gene	Class	Remarks
Alanine	ALA1	II (cytoplasmic and mitochondrial)	cytoplasmic and mitochondrial alanyl-tRNA synthetase; point mutation (*cdc64-1*) causes cell cycle arrest at G_1; lethality of null mutation is functionally complemented by human homolog
Arginine	RRS1	I (cytoplasmic)	proposed to be cytoplasmic but the authentic, nontagged protein is detected in highly purified mitochondria
	MSR1	I (mitochondrial)	
Asparagine	DED81	II (cytoplasmic)	
	SLM5	I (mitochondrial)	probable mitochondrial seryl-tRNA synthetase; mutant displays increased invasive and pseudohyphal growth
Aspartate	DPS1	II (cytoplasmic)	primarily cytoplasmic; homodimeric enzyme that catalyzes the specific aspartylation of $tRNA^{Asp}$; binding to its own mRNA may confer autoregulation
	MSD1	II (mitochondrial)	yeast and bacterial aspartyl-, asparaginyl-, and lysyl-tRNA synthetases contain regions with high sequence similarity, suggesting a common ancestral gene
Cysteine	YNL247w	I (cytoplasmic)	may interact with ribosomes, based on copurification experiments
Glutamate	GUS1, *ERS*	I (cytoplasmic)	GluRS; forms a complex with methionyl-tRNA synthetase (Mes1p) and Arc1p; complex formation increases the catalytic efficiency of both tRNA synthetases and ensures their correct localization to the cytoplasm
	MSE1	I (mitochondrial)	mitochondrial glutamyl-tRNA synthetase, predicted to be palmitoylated
Glutamine	GLN4, QRS	I (cytoplasmic)	N-terminal domain proposed to be involved in enzyme–tRNA interactions
	none	I (mitochondrial)	synthetase missing
Glycine	GRS1	II (cytoplasmic and mitochondrial)	transcription termination factor that may interact with the 3′-end of pre-mRNA to promote 3′-end formation; *GRS2* is probably a pseudogene
Histidine	HTS1	II (cytoplasmic and mitochondrial)	cytoplasmic and mitochondrial histidine tRNA synthetase; encoded by a single nuclear gene that specifies two messages; efficient mitochondrial localization requires both a presequence and an N-terminal sequence
Isoleucine	ILS1	I (cytoplasmic)	target of the G_1-specific inhibitor reveromycin A
	ISM1	I (mitochondrial)	null mutant is deficient in respiratory growth
Leucine	CDC60	I (cytoplasmic)	*CDC60* was first identified as a temperature-sensitive mutant that arrested at START upon shift to the restrictive temperature; the cell cycle arrest of the mutant is probably due to the block in protein synthesis that results from a lack of charged leucyl-tRNA
	NAM2	I (mitochondrial)	Also has a direct role in splicing of several mitochondrial group I introns; indirectly required for mitochondrial genome maintenance
Lysine	KRS1	II (cytoplasmic)	a second lysyl-tRNA synthetase, Msk1p, is localized to mitochondria; both the cytoplasmic and mitochondrial enzymes are required for the import of nuclear encoded $tRNA^{Lys(CUU)}$ into mitochondria (Tarassov, Entelis, and Martin, 1995); the expression of *KRS1* is under general amino acid control and a mutant was initially characterized as a negative regulator of general control of amino acid biosynthesis
	MSK1	II (mitochondrial)	
Methionine	MES1, cMRS	I (cytoplasmic)	forms a complex with glutamyl-tRNA synthetase (Gus1p) and Arc1p, which increases the catalytic efficiency of both tRNA synthetases; also has a role in nuclear export of tRNAs
	MSM1	I (mitochondrial)	MetRS; functions as a monomer in mitochondrial protein synthesis; functions similarly to cytoplasmic MetRS although the cytoplasmic form contains a zinc-binding domain not found in Msm1p
Phenylalanine	FRS1	II (cytoplasmic)	cytoplasmic phenylalanyl-tRNA synthetase β-subunit, forms a tetramer with Frs2p to generate active enzyme; able to hydrolyze mis-aminoacylated $tRNA^{Phe}$, which could contribute to translational quality control
	FRS2		cytoplasmic phenylalanyl-tRNA synthetase α-subunit, forms a tetramer with Frs1p to form active enzyme; evolutionarily distant from mitochondrial phenylalanyl-tRNA synthetase based on protein sequence, but substrate binding is similar
	MSF1	II (mitochondrial)	active as a monomer, unlike the cytoplasmic subunit; similar to the α-subunit of *E. coli* phenylalanyl-tRNA synthetase
Proline	YHR020w	II (cytoplasmic)	based on copurification experiments; has similarity to proline tRNA ligase; essential gene
	AIM10	II (mitochondrial)	protein with similarity to proline tRNA synthetases; nontagged protein is detected in purified mitochondria; null mutant is viable
Serine	SES1	II (cytoplasmic)	displays tRNA-dependent amino acid recognition which enhances discrimination of the serine substrate, interacts with peroxin Pex21p
	SLM5	II (mitochondrial)	see SLM5 above in 'Asparagine'

Table 5.8 (*Continued*)

Amino acid	Gene	Class	Remarks
Threonine	THS1	II (cytoplasmic)	essential cytoplasmic protein
	MST1	I (mitochondrial)	
Tyrosine	TYS1	I (cytoplasmic)	interacts with positions 34 and 35 of the $tRNA^{Tyr}$ anticodon; mutations in human ortholog YARS are associated with Charcot–Marie–Tooth neuropathies
	MSY1	I (mitochondrial)	
Tryptophan	WRS1	I (cytoplasmic)	
	MSW1	I (mitochondrial)	
Valine	VAS1	I (cytoplasmic and mitochondrial)	

required. This allows most of the cERS and cMRS to be sequestered by the Arc1p–ERS–MRS complex. When yeast switches to aerobic metabolism, the activity of Arc1p is drastically reduced, so that both cERS and cMRS are released and ready to be recycled for use in protein synthesis. This implies that Arc1p also coordinates the mitochondrial and cytosolic translation levels in response to the switch in nutritional carbon sources.

5.2.2.3.2 **Rules, Codon Recognition, and Specific tRNA Modification** In a **"tRNomics" approach**, the complete set of tRNAs required in each genome was compared along 19 structural criteria (Grosjean, de Crécy-Lagard, and Marck, 2010). Some **kingdom-specific rules** could be revealed: (i) only eukaryal tRNAs contain the consensus A and B boxes (*cf.* Section 9.2); (ii) in eukaryal $tRNA^{His}$, the G1 residue has to be added post-transcriptionally, whereas in bacteria and Archaea it is encoded in the tDNA; (iii) base pairs 11–24 are always Y11–R24 in eukaryotes, while in bacterial and archeal tRNAs they are always R11–Y24. This approach also permitted to uncover three **anticodon-sparing principles**, of which two apply in all domains of life, while the third is only pertinent to bacteria. The "A34- or G34-sparing strategy" demands that tRNAs with A34 decode codons like NNU as well as codons like NNC, meaning that 46 different anticodons are sufficient to serve 62 codons. The "A34- or G34- *and* C34-sparing strategy" takes advantage of the capability of U34 to pair with any of the four nucleotides at the third codon position, thus sparing codons with C34. This rule reduces the number of different anticodons that can read 62 codons to only 33. The third rule (applying only to bacteria) says that the "A34- or G34- *and* C34-sparing strategy" affords only 26 different anticodons.

Although in most cases the significance of modified nucleotides in tRNAs remains unknown, there is accumulating evidence that modifications in particular positions are required for integrity and full function of a tRNA (Bjork *et al.*, 1999; Huang *et al.*, 2008; Giegé, 2008; Gustilo, Vendeix, and Agris, 2008). Moreover, post-transcriptional modification patterns have revealed other unique features of tRNAs from specific domains of life and full maps of such modifications in all three kingdoms have been established (Grosjean, de Crécy-Lagard, and Marck, 2010). Some of the rules that arose are briefly summarized in Figure 5.23.

5.2.2.3.3 **Recognition of tRNAs in the Protein Biosynthetic Network** The prediction of **identity determinants** is difficult. Either it can be achieved by experimental approaches or may profit from computational comparisons of tRNA features (e.g., McClain, 1993; Ardell, 2010), whereby usually

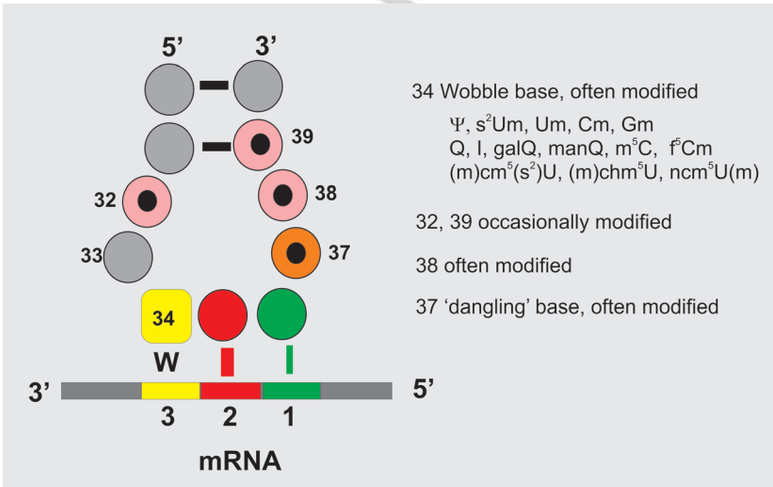


Fig. 5.23 Anticodon–codon pairing in eukaryotes. Numbering follows the standard nomenclature; W, wobble position.

extensive searches among different classes of tRNAs and many organisms are required, and experimental proof of the results is recommended.

As one of the earlier examples, the structure and function of *in vitro* transcribed $tRNA^{Asp}$ variants with inserted conformational features characteristic of yeast $tRNA^{Phe}$, such as the length of the variable region or the arrangement of the conserved residues in the D-loop, have been investigated (Perret *et al.*, 1992). Although they exhibited significant conformational alterations as revealed by Pb^{2+} treatment, these variants were still efficiently aspartylated by yeast aspartyl-tRNA synthetase. Thus, this synthetase can accommodate a variety of tRNA conformers. In a second series of variants, the identity determinants of yeast $tRNA^{Phe}$ were transplanted into the previous structural variants of $tRNA^{Asp}$. The phenylalanine acceptance of these variants improved with increasing the number of structural characteristics of $tRNA^{Phe}$, suggesting that phenylalanyl-tRNA synthetase became sensitive to the conformational frame embedding the cognate identity nucleotides. These results contrasted with the efficient transplantation of $tRNA^{Asp}$ identity elements into yeast $tRNA^{Phe}$. The authors concluded that synthetases respond differently to the detailed conformation of their tRNA substrates. Efficient aminoacylation is not only dependent on the presence of the set of identity nucleotides, but also on a precise conformation of the tRNA.

Eukaryotic serine-specific tRNAs are characterized by possessing a large variable (or extra) loop; the sequence CCCG in its stem plus G73 form the serine identity element. For leucine identity, the sequence of the anticodon stem and bases of the anticodon loop (positions 27–35) are responsible. Recently, the universal identity determinant of alanine-specific tRNAs has been recognized as a single base pair (G3 : U70) (Reebe *et al.*, 2008).

The problem of identity determinants and antideterminants gets more complex in the cases of the two eukaryotic **methionine-specific tRNAs** – the initiator and the elongator (Figure 5.24).

As can directly be inferred from a comparison of the clover-leaf structures, the sequences deviate from each other substantially, although both share the same sequence in the anticodon triplet (designed to read the codon AUG) and the anticodon loop (Kolitz and Lorsch, 2010). Since the initiator tRNA in eukaryotes is not marked by formylation as in bacteria, the discrimination of the two distinct functions of the two methionine tRNAs completely resides in the residual structure (i.e., that each type of methionyl-tRNA has to be restricted to its separate function). Nonetheless, both tRNAs have to be recognized and charged by the *same* enzyme, methionyl-tRNA synthase. After this, the fates of the charged tRNAs diverge – the initiator binds eIF2-GTP and must not bind eEF1A. A major determinant for exclusion of the initiator of binding eEF1A is the A1–U72 base pair conserved in eukaryotic initiator tRNAs. Contrary to bacteria, the T-loop in eukaryotic initiator tRNAs deviates from that of the elongator: instead of having a "normal" TψC-loop, the initiator lacks the T54–Ψ55 and has an A54–U55. The T54 acts as an important determinant for elongator in eEF1A recognition.

The initiator tRNA is thought to bind directly to the P-site of the small subunit of the ribosome and to have a critical role in recognizing the start codon in mRNA. In this process, both the initiation factors and the tRNA cooperate. In this context, the overall structure of the initiator appears to be of high relevance, mainly the differences seen in the D-loop and the T-loop. Eukaryotic initiators lack nucleotide 17, which is present in all other tRNAs; the D-loop contains A20, also found in the yeast elongator (while most other tRNAs contain a D in this position). The T-loop contains A54, exclusively present in eukaryotic initiators. Also, A60 is unusual in the initiator, as all eukaryotic elongators have a pyrimidine in this position. Obviously, hydrogen bonding

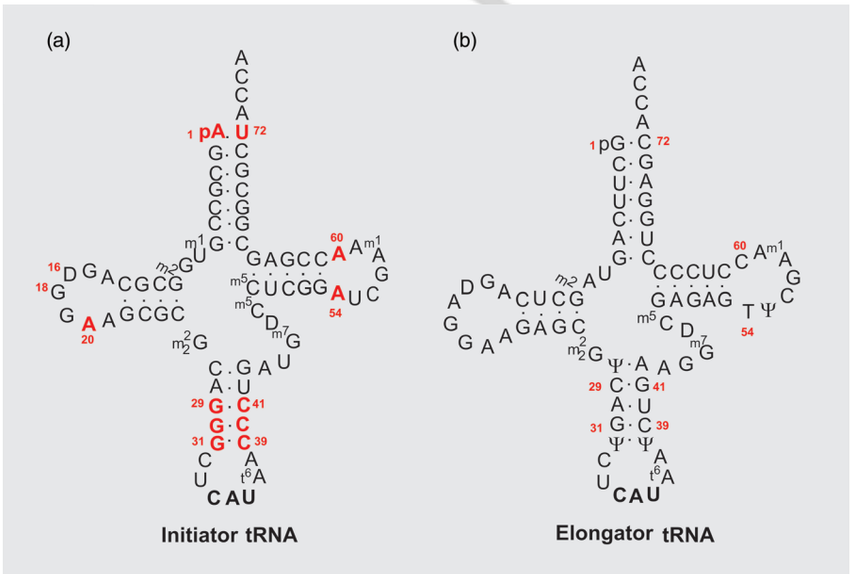


Fig. 5.24 Methionine-specific tRNAs from *S. cerevisiae*.

thus induced between A20 in the D-loop and G57, A59, and A60 in the T-loop create a stronger connection between these parts, which are less strong in the elongator.

The presence of three consecutive G–C base pairs in the anticodon stem (which with a few exceptions are present in initiator tRNAs in all three domains of life) fostered the idea that the anticodon stem–loop might have a unique conformation. Indeed, the crystal structure of initiator revealed a spatial, more compact conformation, which is further supported by interactions between C32 and A38, and a base triplet formed between A37, G29, and C41 (Barraud *et al.*, 2008). In all, the yeast initiator tRNA carries 11 nucleotide modifications, of which m1A58 plays an eminent role – it is responsible for the stability of the initiator tRNA. Molecules lacking this modification are subjected to degradation by the nuclear Trf4p/Rpr6p surveillance pathway (Kadaba *et al.*, 2004).

5.3 Yeast Ribosomes: Components, Genes, and Maturation

5.3.1 Historical Overview

Ribosomes – the protein factories. Undoubtedly, the pioneering work on ribosomes, rRNAs, and ribosomal proteins was done in the *E. coli* system. In the early days of molecular biology, there was no need to fall back upon eukaryotic organisms, since appropriate material was abundantly available from bacterial sources. Also, detailed work on ribosome structure and function was guided by investigations into bacterial ribosomes. Only when it became apparent through the work of Hartwell and colleagues that the constituents of eukaryotic ribosomes differed in several aspects from their bacterial counterparts, such as in type of RNA or number of ribosomal proteins (Helser and McLaughlin, 1975; Kaback *et al.*, 1976; Kaback and Halvorson, 1977), as well as in size, arrangement (Philippsen *et al.*, 1978; Kaback and Davidson, 1980), and expression of the rRNA genes, did yeast become a player in this field.

The earliest research on yeast ribosomes stems from the work of J.R. Warner (Warner, 1971), when he investigated the **assembly of ribosomes** in yeast, followed by numerous publications from his laboratory on the ribosome field (Warner, 2001; Warner *et al.*, 1973). When it became feasible to study gene regulation in yeast, the genes for rRNAs and ribosomal proteins were of particular interest. A hierarchy of elements regulating the synthesis of yeast ribosomal proteins (Kief and Warner, 1981; Kim and Warner, 1983; Mitra and Warner, 1984) and later the effects of nutritional control on ribosome synthesis were described (Mizuta *et al.*, 1998; Li, Nierras, and Warner, 1999). Reb1p, a key regulator of yeast ribosome synthesis, was described in 1990 (Ju, Morrow, and Warner, 1990) and a new regulator was discovered in 2004 (Wade, Hall, and Struhl, 2004). Description of promoter and terminator elements for rRNA synthesis as well as the *trans*-regulatory control proteins began in 1984 and has been continued to the present (Elion and Warner, 1984; Morrow, Ju, and Warner, 1990; Lang *et al.*, 1994; Wang and Warner, 1998; Zhao, Sohn, and Warner, 2003). The three-dimensional structure of the yeast ribosome has been established (Verschoor *et al.*, 1998).

Measured in terms of sequence determination, **rRNA genes and ribosomal protein genes** from yeast were somewhat behind what had been established in *E. coli*. Sequences of the small rRNAs from *E. coli* were established in 1967 (Brownlee, Sanger, and Barrell, 1967), and those of the large rRNA genes in 1978 and 1980, respectively (Brosius *et al.*, 1978; Brosius, Dull, and Noller, 1980). In comparison, the sequences of the yeast rRNA genes were solved a few years later: 5S (Valenzuela *et al.*, 1977), 5.8S (Rubin, 1973), 18S (Rubtsov *et al.*, 1980), and 25S (Veldman *et al.*, 1981). The first DNA sequences containing yeast ribosomal protein genes were isolated in 1979 (Woolford, Hereford, and Rosbash, 1979) and later on these genes were characterized (Fried *et al.*, 1981).

5.3.2 Ribosomal Components

Composites of RNAs and proteins. About 60% of total cellular transcription is committed to that of the rRNA genes by RNA polymerase I, which comprise about 10% of the entire genome. While all mRNAs together only comprise about 5% of total cellular RNA, it can be estimated that expression of the ribosomal protein genes represents about half of the cell's transcriptional capacity by RNA polymerase II, although the ribosomal protein genes occupy only 2% of the yeast genome. RNA polymerase II is also responsible for the production of the majority of the small nucleolar RNAs, which are collectively involved in maturation of the ribosome.

5.3.2.1 Ribosomal RNAs

In yeast, the transcriptional units for the **four rRNA genes** are organized in tandemly **repeated units** of 9.1 kb length. About 120 copies of these units are localized in a coherent 1–2 Mb region on the right arm of chromosome XII. We will elaborate in Chapter 9 on how transcription of the rRNA genes is accomplished. Note that three of the rRNAs (18S of 1798 nucleotides, 5.8S of 158 nucleotides, and 25S of 3392 nucleotides, in this order) are transcribed into a common precursor (35S RNA), which still contains external and internal spacer sequences, whereas the 5S (of 121 nucleotides) RNA gene is separately transcribed by RNA polymerase III (Figure 5.25).

The rDNA of *S. cerevisiae* is encoded by the so-called *RDN1* locus. Transcription starts in the 5′-ETS and terminates in the 3′-ETS. The majority of transcripts terminate at a terminator 93 bp downstream of the 3′-end of 25S rRNA, while a minority terminate at a site 211–250 nucleotides

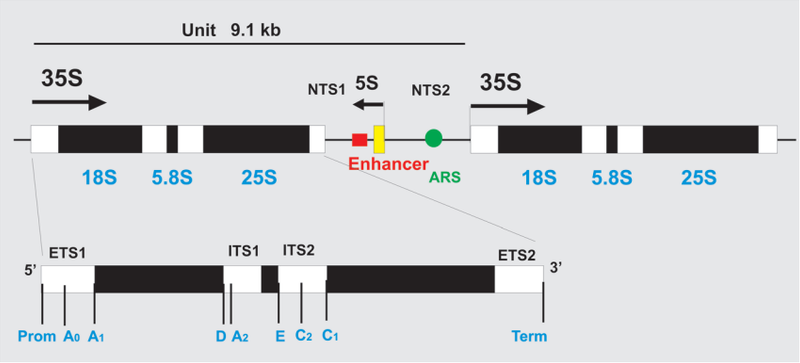


Fig. 5.25 Transcriptional units for yeast rRNAs. NTS, nontranscribed spacer; ETS, external transcribed spacer; ITS, internal transcribed spacer; Prom, promoter; Term, terminator.

downstream. The 5S rRNA is transcribed separately, and on the opposite strand, by RNA polymerase III.

The systematic sequencing of the yeast genome included only two of the rDNA repeats, but each of the two annotated repeats is represented by several locus-specific entries. RDN37-1 and RDN37-2 contain the sequences of the primary 35S transcripts of the two repeats, while RDN25-1 and RDN25-2, RDN18-1 and RDN18-2, and RDN58-1 and RDN58-2 represent the 25S, 18S, and 5.8S rRNAs encoded by these transcripts, respectively. Also, the sequences for each of the four nontranscribed spacer (NTS), external transcribed spacer (ETS), and internal transcribed spacer (ITS) sequences have been listed. The story of the 5S rRNA genes associated with these two large rDNA units is much more complicated – two copies reside within the *RDN1* locus, but only one reflects the complete sequence of intact 5S RNA (as indicate in Figure 5.25); further copies of the 5S RNA genes are located distal to RDN1 in a 3.6-kb region towards the telomere and revealed variant 5S RNA sequences.

5.3.2.2 Ribosomal Proteins

The genes encoding the 78 different **ribosomal proteins**, 32 for the small 40S and 42 for the large 60S ribosomal subunits, respectively, are scattered throughout the genome; as a rule, two identical or nearly identical genes exist for each of these proteins. The corresponding genes and proteins are accurately listed in the yeast databases (e.g., www.yeastgenome.org). Four of the ribosomal proteins are components of the ribosomal stalk, which is involved in the interaction of translational elongation factors with the ribosome and regulated by phosphorylation; two of the ribosomal proteins are fusion proteins that upon cleavage yield ubiquitin and one component of the small and the large subunits, respectively. Ubiquitin may facilitate the assembly of the ribosomal protein into ribosomes.

5.3.3 Components and Pathways of Yeast Ribosome Maturation

Ribosome assembly is an extremely complex process. The assembly of the mature particles needs a lot of different steps occurring in different parts of the cell (overviews: Warner, 2001; Fatica and Tollervey, 2002; Henras *et al.*, 2008). The synthesis of the ribosomal **35S precursor** RNA occurs in the yeast nucleolar organizer, while the ribosomal proteins and additional nonribosomal proteins (about 200 needed in the maturation process) are synthesized in the cytosol and imported into the nucleolus. Before the single rRNA species are cleaved out from the precursor, extensive chemical **modifications** have to be carried out by modifying enzymes. These include methylation of 2′-OH positions on nucleotide ribose moieties at various positions, isomerization of particular uridine residues into pseudouridine residues, and methylation on a few adenine nucleotides. The functions of these modifications are not understood in detail; presumably they assist later in correct cleavage and folding of the rRNA. The exact positions of the modifications (with one exception) are determined with the aid of **"guide" RNAs**, the so-called small nucleolar RNAs (snoRNAs), which are manufactured and recycled in the nucleolus. These small RNAs locate themselves to the specific regions of the rRNA precursor by complementary base-pairing, thereby bringing the associated RNA-modifying enzymes (packaged into small RNP particles (snoRNPs)) to the appropriate positions. Other guide RNAs are designed to recruit endo- and exonucleolytic enzymes to positions in which sequences have to be cut out from the precursor molecules in order to generate the mature 18S, 5.8S, and 25S RNA moieties.

Helpers in assembly. There are 76 stable yeast **snoRNAs** that serve as constituents of snoRNPs. (An overview on all snoRNAs in yeast can be obtained from http://www.yeastgenome.org). Although the majority of the snoRNPs are involved in rRNA processing, some of them participate in the processing of other RNAs, such as tRNAs or the telomerase RNA TCL1. Based on conserved sequence elements, the snoRNAs can be divided into three classes:

i) **Box C/D snoRNAs** contain one or more sequences, from 10 to 22 nucleotides long, of perfect complementarity to the sequence of their target RNA molecule, most often either the 18S or 25S rRNAs, which are flanked by two short conserved sequence elements, called boxes C and D, located near the 5′- or 3′-end of the snoRNA, respectively. Each box C/D snoRNA is

bound by four evolutionarily conserved proteins to form a box C/D-type small nucleolar RNP complex (snoRNP): Nop1p (methyltransferase, the homolog of vertebrate fibrillarin), Nop58p, Nop56p, and Snu13p.

ii) **Box H/ACA snoRNAs** typically adopt a conserved secondary structure consisting of two hairpins connected by a hinge region that contains the box H sequence motif; the second hairpin is followed by the sequence motif "ACA," which is always three nucleotides upstream from the mature 3′-end of the snoRNA. The site(s) of pseudouridylation is specified by internal loops, found in one or both of the hairpins, containing a stretch of from 9 to 13 nucleotides complementary to the target RNA. Each H/ACA snoRNA associates with a set of conserved proteins: Cbf5p (the pseudouridine synthase catalytic subunit), Gar1p, Nhp2p, and Nop10p to form a H/ACA-type snoRNP.

iii) **MRP snoRNA.** NME1 is a sole representative of this type.

The snoRNPs occur in all eukaryotes and even in Archaea, which documents their ancient descendence and evolutionary high conservation.

The **genomic organization** of the box C/D snoRNAs in *S. cerevisiae* is noteworthy because of its variability. Some of the genes are encoded within the introns of protein coding genes, as is also the case for vertebrate snoRNAs. Other such genes are found in polycistronic arrays, containing from two to seven genes – an organization that is common for plant snoRNAs. Finally, there are also independently transcribed monocistronic box C/D snoRNA genes. The genomic organization of the box H/ACA snoRNAs is less variable, as no such genes are found within polycistronic transcripts, but almost all of them are monocistronic genes; only a couple are found within the introns of protein-coding genes.

While most of the snoRNAs are not essential and involved in RNA nucleotide modification, a few, including members of each of the three families, are required for **endonucleolytic cleavage** steps in the processing to convert the primary rRNA transcript into the mature 18S, 5.8S, and 25S rRNA molecules.

Scheduling the process. To date, the **90S preribosome complex** (Figure 5.26) is described as corresponding to the earliest detectable rRNA processing and ribosome assembly complex. The 90S preribosome has also been characterized biochemically and shown to contain around 35 nonribosomal components, including proteins associated with snoRNP U3 (e.g., Nop56p, Nop58p, Sof1p, Rrp9p, Dhr1p, Imp3p, Imp4p, and Mpp10p) and many other factors required for 18S rRNA synthesis (Grandi *et al.*, 2002). Further, a number of U3-containing early ribosome assembly and rRNA processing complexes have been identified that contain the 35S pre-rRNA transcript and have overlapping but not identical protein compositions (Granneman and Baserga, 2004). Of particular interest is the so-called small subunit (SSU) **processome complex**, which is a large (greater than 2 MDa) complex composed of the U3 snoRNA and associated so-called Utp proteins that cotranscriptionally assembles at the 5′-end of the nascent pre-rRNA (Dragon *et al.*, 2002). Both the 90S preribosome and the SSU processome (or processosome) complexes contain ribosomal proteins, primarily of the small 40S ribosomal subunit, and nonribosomal proteins that are involved in rRNA processing and assembly of the small subunit.

The **U3 snoRNA** is one of the most abundant RNA molecules in *S. cerevisiae*, present in about 400–1000 copies per cell. U3 is a box C/D molecule encoded by two genes, *SNR17A* and *SNR17B*, both of which contain an intron with an atypical branch point sequence. Both U3 genes are transcribed, 328 nucleotides long, and 96% identical in the region of the mature RNA. U3 from *S. cerevisiae* is over 100 nucleotides longer than U3 from most other eukaryotes (e.g., human, rat, or *Dictyostelium*), but shares conserved primary and secondary structure elements, including perfect complementarity to a conserved sequence within the 5′-ETS of the primary rRNA transcript and to three highly conserved sequences within the 18S rRNA, which form the conserved pseudoknot found at the core of all small subunit rRNAs.

Following the assembly of 90S preribosomes, three early endonucleolytic cleavages in the 35S pre-rRNA by Pwp2p (Utp1p) endonuclease (at sites A_0, A_1, and A_2 of the 35S rRNA precursor; Figure 5.25) generate the 20S and $27SA_2$ pre-rRNAs – the precursors to 18S and 25S/5.8S rRNA, respectively. The endonuclease contains eight WD repeats, and its deletion leads to defects in cell cycle and bud morphogenesis. Consequently, these cleavages initiate 40S and 60S subunit formation. Two box C/D snoRNPs, U3 and U14 (produced by *SNR128*), and two box H/ACA snoRNPs, snR30 and snR10, are required for cleavage of the primary rRNA transcript. Depletion of U3, U14, or snR30 results in depletion of the 18S rRNA, and complete lack of any one of these snoRNAs is lethal. U14 and snR10 are involved in both endonucleolytic cleavage steps and in targeting RNA modification reactions. In addition, RNase Mrp1p is involved in endonucleolytic cleavage to produce the mature 5.8S rRNA molecule.

Processing of the larger preribosome yields **pre-40S**, whereby most of the auxiliary proteins are removed (Schäfer *et al.*, 2003; Granneman *et al.*, 2010); only a few of the 90S factors (e.g., Enp1p) were shown to remain associated with 20S pre-rRNA (Grandi *et al.*, 2002). A single cleavage reaction converts 20S pre-rRNA into mature 18S rRNA – a process that requires at least RNase Rrp10p (Rio1p) (Vanrobays *et al.*, 2001). The 20S pre-rRNA also undergoes adenine dimethylation, catalyzed by the Dim1p dimethylase, at two positions close to its 3′-end. Nuclear fractionation experiments indicated that both the modification and 3′ cleavage of the 20S pre-rRNA occur after export to the cytoplasm. Additional factors participating in 40S maturation are: Nob1p, the PIN-domain endonuclease that cleaves site D at the 3′-end of 18S rRNA (Fatica *et al.*, 2003); Tsr1p, a small GTPase, required as a cofactor in cleavage of the 20S pre-rRNA at site

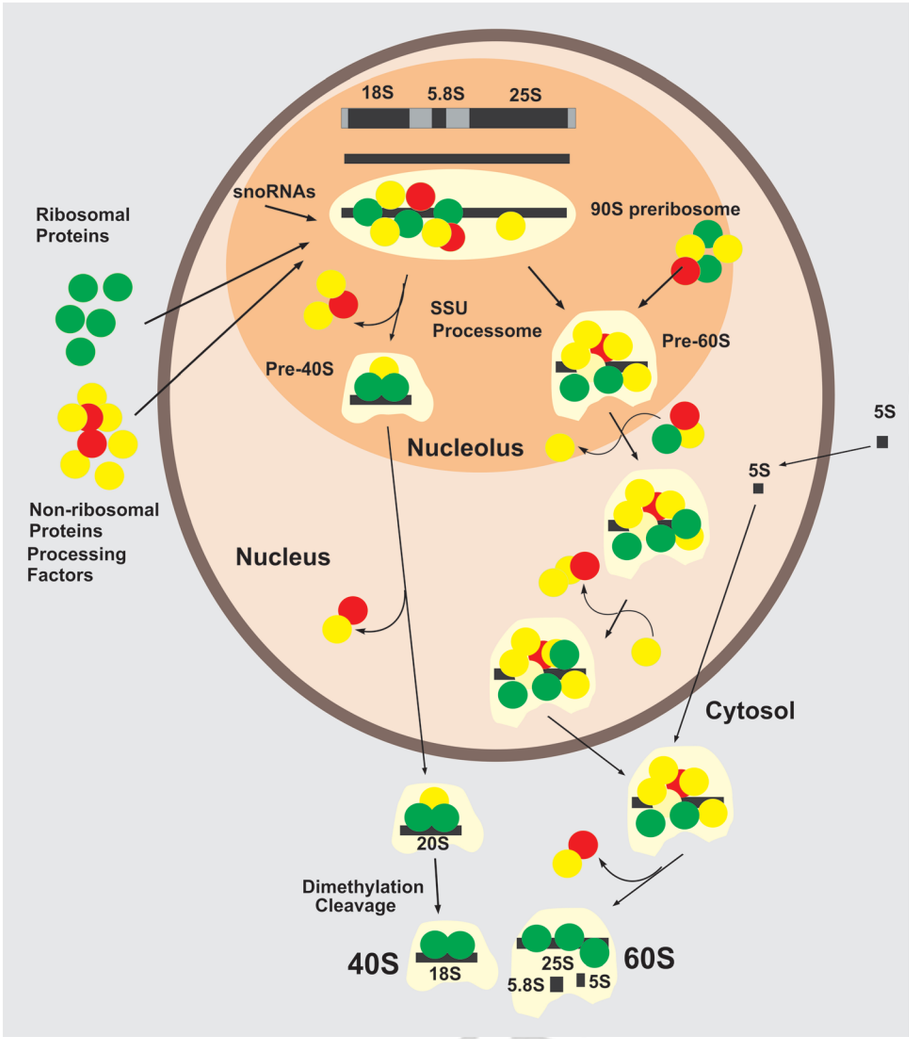


Fig. 5.26 Processing of yeast rRNAs and proteins.

D (Gelperin *et al.*, 2001); and other late-acting 40S synthesis factors such as kinase Rio2p and Prp43p (Bohnsack *et al.*, 2009). Nuclear export of 40S preribosomes requires Xpo1p and nucleoporins of the Nup82p complex (Moy and Silver, 2002), but which pre-40S components carry nuclear export signals (NESs) is unknown to date.

Several distinct **pre-60S** ribosomal particles were identified that differ in their content of associated nonribosomal proteins and pre-rRNA species on the pathway of 25 rRNA and 5.8S rRNA synthesis (Bassler *et al.*, 2001; Saveanu *et al.*, 2001; Nissan *et al.*, 2002). Compared to pre-40S, a different set of nonribosomal proteins assembles onto the 27S pre-rRNAs to generate pre-60S subunits. Processing of the 27S pre-rRNAs into 25S and 5.8S rRNA particles occurs in several distinct pre-60S intermediates (one of which was termed the 66S precursor), which pass from the nucleolus via the nucleoplasm to the nuclear periphery (Bassler *et al.*, 2001; Saveanu *et al.*, 2001; Fatica and Tollervey, 2002; Nissan *et al.*, 2002). Finally, nuclear export of pre-60S particles requires Nmd3p, an adaptor protein that carries a NES and serves to couple the large subunit protein Rpl10p to the nuclear export receptor Xpo1p/Crm1/exportin-1 (Gadal *et al.*, 2001a; Gadal *et al.*, 2001b).

During 35S pre-rRNA maturation, the spacers are removed in an ordered series of exo- and endonucleolytic processing reactions that involve many proteins and snoRNAs. To date, 14 putative RNA helicases, most of which are essential for cell viability, have been implicated in ribosome synthesis. They belong to the large DEAD/DExH-box family in yeast (*cf.* Section 6.5.1).

5.4 Messenger RNAs

5.4.1 First Approaches to the Structure of Yeast mRNAs

How do yeast mRNAs compare to their mammalian counterparts? Studies on yeast mRNA started around 1969–1970 in

the laboratory of Lee Hartwell with the participation of Warner and McLaughlin (Hutchison, Hartwell, and McLaughlin, 1969). As soon as it became known that mammalian mRNA is **polyadenylated** (Edmonds and Caramela, 1969; Edmonds, Vaughan, and Nakazato, 1971), McLaughlin *et al.* (1973) were able to show that mRNAs from yeast contain polyadenylic acid sequences of around 50 nucleotides in length at their 3′-ends and a few years later it was established that the 5′-termini of mRNA from yeast are blocked by methylated nucleotides (Sripati, Groner, and Warner, 1976). Interestingly, even yeast histone mRNA was found to contain 3′-poly(A) sequences (Fahrner, Yarger, and Hereford, 1980).

Yeast mRNAs for specific proteins were isolated and characterized only later (Holland, Hager, and Rutter, 1977). For example, mRNAs for glycolytic enzymes were identified in 1978 (Holland and Holland, 1978) and ribosomal protein genes in 1980 (Bollen *et al.*, 1980). A fact mentioned in Michael Smith's Nobel Lecture (Smith, 1993) is that his first cooperation for applying his newly developed approaches of using **synthetic deoxyribooligonucleotides for monitoring gene isolation** involved the people working on the yeast iso-1-cytochrome *c* gene – the laboratories of Fred Sherman and Benjamin D. Hall. Stewart and Sherman (1974) had identified frameshift mutations by sequence changes in iso-1-cytochrome *c*. This led to the enzymatic synthesis of oligonucleotides of defined sequence for identifying this gene (Gillam *et al.*, 1977), its isolation, and sequence determination (Montgomery *et al.*, 1978; Smith *et al.*, 1979). The sequence of the iso-1-cytochrome *c* (*CYC1*) mRNA was also determined (Boss *et al.*, 1981), as well as its 5′-end positioned by *in vitro* mutagenesis, using synthetic duplexes with random mismatch base pairs (McNeil and Smith, 1985). Later, Guarente and collaborators studied the regulation of *CYC1* (Guarente, 1987; Olesen, Hahn, and Guarente, 1987) and *CYT1* (cytochrome c_1) by heme via the HAP complex (Schneider and Guarente, 1991).

Both the aforementioned techniques were applied to the SUP4 $tRNA^{Tyr}$ locus (Koski *et al.*, 1980; Kurjan *et al.*, 1980). Thus, these initial approaches made clear that "synthetic DNA" became an invaluable tool for many applications – as a probe for gene isolation, in direct sequencing of double-stranded DNA by the enzymatic method of Sanger's laboratory (Sanger, Nicklen, and Coulson, 1977) using synthetic oligonucleotide primers, for the precise identification of point mutations produced by classical genetic techniques at a given locus, or in the development of oligonucleotide-directed mutagenesis.

The interest in isolating **specific mRNAs** from yeast probably faded once the cloning of specific yeast genes became feasible. The lack of large introns in yeast genes and an average size of yeast genes of some kilobases meant a huge advantage in the cloning strategies over genes from higher eukaryotes, where introns could be manifold and of considerable length.

5.4.2 Introns and Processing of pre-mRNA

The yeast splicing machinery becomes a paradigm. It was in 1977 that the occurrence of **introns** in mammalian genes was pinned down and that splicing was detected as the decisive step in maturation of pre-mRNA to mature mRNA by the Nobel Prize winners of 1993, Roberts (1993) and Sharp (1993), not to forget the merits of others (Berget, Moore, and Sharp, 1977; Breathnach, Mandel, and Chambon, 1977; Gelinas and Roberts, 1977; Jeffreys and Flavell, 1977; Tilghman *et al.*, 1978). Only 3 years later, the actin gene from yeast was shown by Gallwitz and Sures to possess an intron sequence near its 5′-end (Gallwitz and Sures, 1980).

Although it became clear much later that only 4–5% of the yeast genes possess introns, the sophisticated **splicing machinery** of eukaryotic organisms has been fully retained in yeast. Finally, more than 100 different genes encode products important for pre-mRNA splicing, comprising about 2% of the total yeast genome. In fact, yeast has served as a model system that has substantially contributed to fully disentangle the "splice cycle" genetically and biochemically, mainly initiated by the work of C. Guthrie, J. Abelson, J. Beggs, and their collaborators (Staley and Guthrie, 1998; Stevens and Abelson, 2002). Nonetheless, details of the splicing mechanism are still under study to date (Rader and Guthrie, 2002; Perriman *et al.*, 2003; Silverman *et al.*, 2004; Grainger and Beggs, 2005).

In 1983, Langford and Gallwitz (1983) described a (unique) intron-contained sequence in yeast required for splicing – the so-called **branch point**, which was also observed in polyadenylated RNA from other sources (Wallace and Edmonds, 1983). In yeast, this site was identified as a particular A residue within the (unique) intron sequence TACTAAC (Langford *et al.*, 1984). In the same year, several groups were able to show that lariat structures are the *in vivo* intermediates of the splicing process, similarly occurring in yeast and in mammalian systems (Padgett, Hardy, and Sharp, 1983; Domdey *et al.*, 1984; Grabowski, Padgett, and Sharp, 1984; Keller, 1984; Krainer *et al.*, 1984; Padgett *et al.*, 1984).

Biochemically, the branch site could be defined as a 2′/3′-ester bond (Konarska *et al.*, 1985). Shortly before, in 1981, it had been recognized by Breathnach and Chambon (1981) that there was a limited set of conserved sequences (preferably 5′-GU . . . AG-3′) at each **intron boundary**, and these consensus sequences were found to be common for vertebrate, plant, and yeast cells (Padgett *et al.*, 1986).

From then on, several groups were engaged in characterizing the cellular components involved in the splicing process and in elaborating the detailed mechanism of this process. In the end, it turned out that mRNA processing followed similar routes in yeast and in higher eukaryotes. The first functionally important components surmised to be involved in splicing were the snRNAs (Ohshima *et al.*, 1981),

which are ubiquitous and had been found in all organisms from bacteria to humans, and in many viruses. While the snRNAs in higher eukaryotes are encoded by up to 100 gene copies each, the laboratory of C. Guthrie detected that yeast contains five **snRNAs** (U1, U2, U4, U5, and U6 snRNAs), each encoded by a *single* copy of an essential gene (Wise *et al.*, 1983). The mutual interactions of snRNAs and their interaction with pre-mRNA, as well as the interdependence of particular splicing steps with particular snRNAs, were studied in detail. The fact that the intermediate state, consisting of two RNAs, was efficiently converted to the final products strongly suggested that these RNAs remain bound in a complex and, given the importance of the snRNAs in splicing, suggested the existence of a "splice cycle" and finally led to the eminently important discovery of the **spliceosome** (Brody and Abelson, 1985; Grabowski, Seiler, and Sharp, 1985). Moreover, the spliceosome was recognized as a particle (much like a ribosome) in which the RNA components were associated with a number of proteins forming stable cellular RNA–protein complexes (Konarska and Sharp, 1987).

The single steps in the **spliceosome cycle**, where particular PRP proteins are required, are consistent with the cycle as defined by kinetic and biochemical methods. Most transitions between specific forms of the spliceosome require one or more specific proteins. Furthermore, a number of PRP mutants were shown to be defective in splicing because of their inability to reassemble snRNPs for further splicing. Thus, both genetic and biochemical results proved that the spliceosome cycle is the process responsible for excision of introns from split genes.

The nearly 100 different proteins shown to cooperate in splicing belong to various types, such as zinc finger proteins, small G-proteins, and ATP-dependent **RNA helicases** of the DEAD- or DExH-box families (Ruby and Abelson, 1988; Schwer and Guthrie, 1991; Strauss and Guthrie, 1991; Madhani and Guthrie, 1992; McPheeters and Abelson, 1992; Sawa and Abelson, 1992). Although the basic mechanisms of pre-mRNA splicing had been resolved in about 15 years from the discovery of spliced genes (Guthrie and Patterson, 1988; Steitz *et al.*, 1988; Guthrie, 1991), in a sort of competition between yeast and higher eukaryotes, the aspects of alternative splicing and *trans*-splicing had to await their resolution with the aid of organisms other than yeast, since these routes scarcely exist in yeast. Furthermore, yeast could contribute only little to solve questions about the evolution of introns and exons.

A vast amount of mainly biochemical data led to a consensus view of an ordered pathway of **spliceosome assembly** that is described in Figure 5.27. The major reactions in pre-mRNA splicing are two *trans*-esterifications that occur in the highly dynamic spliceosome complex. The snRNA–protein **(snRNP) complexes**, known as U1, U2, U4, U5, and U6 snRNPs, are key players. These snRNPs are each composed of the respective U RNAs and a number of small proteins, and in some cases harbor particular splicing factors. These latter are omitted in the overview presented in Table 5.9, but indicated in Table 5.10 and in Figure 5.28. In the biogenesis of the snRNPs, the so-called Sm ring complex is required; it has additional functions during splicing and remains associated with the snRNA as part of the core of each snRNP. It is not yet clear whether the snRNAs are exported to the cytoplasm for assembly into snRNP complexes (as occurs in mammalian cells) or whether the snRNA remains in the nucleus and the Sm ring complex is imported into the nucleus to bind to the snRNA.

U1 is the first snRNP to associate with pre-mRNA, interacting with the 5′-splice site. The U1 snRNA becomes base-

Table 5.9 Composition of spliceosomal small RNP particles.

Particle	RNA	Splice factor	Associated	Common
U1 RNP	U1 snRNA	Prp9p	Mud1p, Mud2p, Snp1p, Snu56p, Snu71p, Nam8p, Yhc1p	Prp5p (bridges U1 and U2)
	snR19	Prp39p Prp40p Prp42p		
U2 RNP	U2 snRNA	Prp11p	Hsh155p, Lsr1p,Lea1p, Ist3p; Msl1p, Cus2p, Ysf3p, Rds3p, Rse1p	
U4 RNP	U4 snRNA	Prp8p	Snu23p	Prp3p, Prp4p, Prp6p, Prp24p, Prp31p are common to the U4, U6, and U5 RNPs
	snR6 snR14			
U6 RNP	U6 snRNA	Prp8p		
U5 RNP	U5 snRNA	Prp8p	Snu23p, Aar2p, Lin1p, Snu114 (GTPase)	
	snR7-S	Prp18p	Slu1p, Slu2p	
	snR14			
NTC	no snRNA	Prp19p Prp46p/ Ntc20p	Cef1p, Cwc2p, Clf1p, Snt309p, Syf1p, Syf2p	
Sm ring complex			Sm heptamer: SmB1, SmD1, SmD2, SmD3, SmE1, SmX2, SmX3	

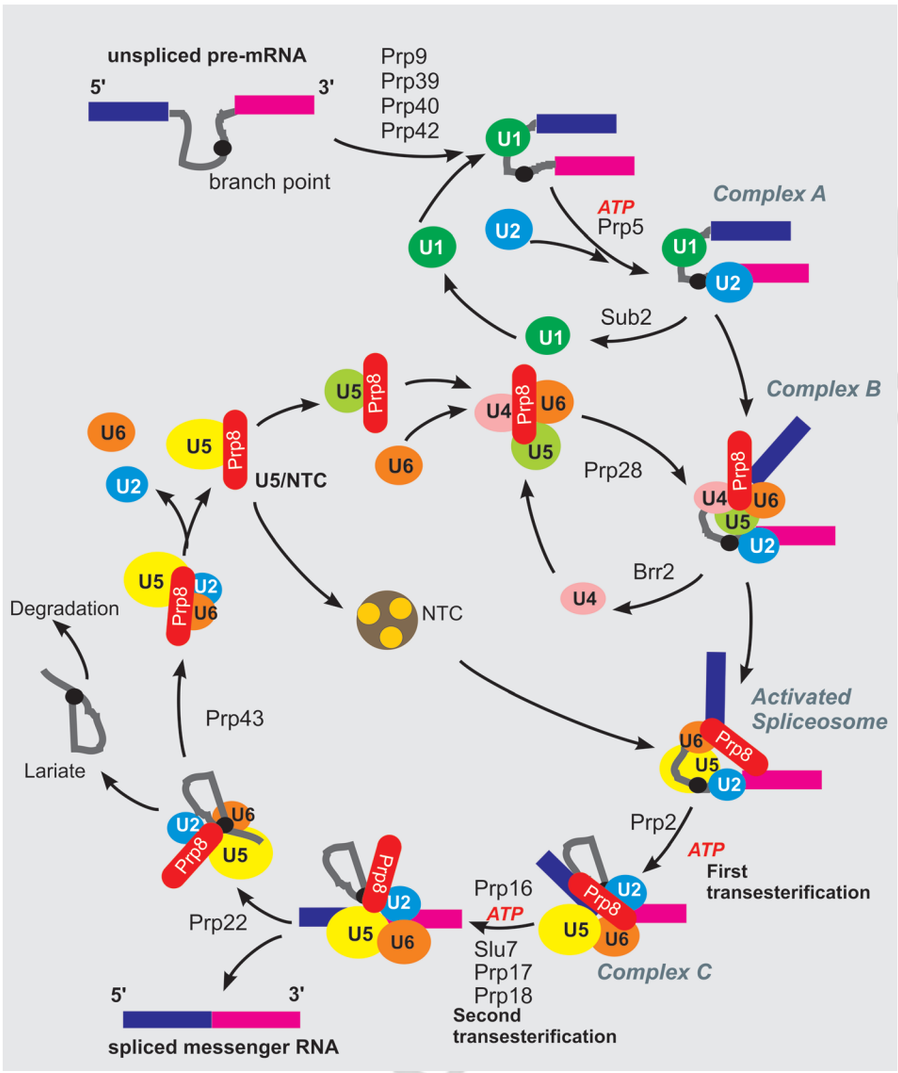


Fig. 5.27 The splice cycle in yeast. Note that the U5 snRNP is depicted in yellow or green depending on its composition. NTC, nineteen complex.

paired to the 5′-end of the intron and a complex forms with the aid of four splicing factors (Prp9p, Prp39p, Prp40p, and Prp42p), committing the pre-mRNA to the splicing pathway. Next, the U2 snRNP is guided to the branchpoint region of the intron by the ATP-dependent helicase Prp5p to form a prespliceosome, also called "complex A." As the U4 and U6 snRNAs share extensive sequence complementarity, they can easily base-pair and form a dimeric U4/U6 snRNP. The U4/U6 complex interacts with U5 snRNP to build a U5·U4/U6 trimeric snRNP, which then associates with the prespliceosome to form "complex B." This reaction is catalyzed by the essential splicing factor Prp8p, which is the largest protein within this machinery (2413 amino acids in length). Prp8p has been shown to have the potential of interacting with many components in the splice cycle and corresponding complexes have been identified (review: Grainger and Beggs, 2005), such as complexes with U5 snRNP, U6 snRNP, pre-mRNA, several reaction intermediates, or the excised intron. Therefore, Prp8p can be viewed as a central turning point of the splicing procedure. Delivery of the U5·U4/U6/Prp8 complex to the prespliceosome occurs by Prp28p.

Formation of the catalytically **competent spliceosome** ("complex C") requires an intricate series of protein and RNA rearrangements, some of which are catalyzed by RNA-dependent NTPases/RNA unwindases: Brr2p, Prp2p, Prp16p, Prp22p, Prp43p, and Sub2p (De la Cruz, Kressler, and Linder, 1999). The concurrent unwinding of the U1 snRNA/5′-splice site and U4/U6 RNA helices is promoted by the U5 snRNP helicases, Prp28p and Brr2p, respectively. The U6 snRNA then base-pairs with the 5′-splice site and with U2 snRNA, to form part of the catalytic center of the spliceosome (Staley and Guthrie, 1998; Staley and Guthrie, 1999). During or after the dissociation of U4 snRNP, the **nineteen complex (NTC)** associates with the assembling active spliceosome in order to stabilize the interactions between the U5 and U6 snRNAs, and remains with U5 there for the consecutive steps. NTC is composed of splicing factor Prp19p and eight other

Table 5.10 Function of spliceosomal proteins.

Factor	Family	Function
Brr2p	Ski2 helicase	unwindase
Prp2p	DExD/H-box	spliceosome activation before first transesterification step
Prp3p, Prp4p, Prp6p, Prp31p		components of U4/U6-U5
Prp5p	DEAD-box	bridges U1 and U2
Prp8p		component of U4/U6-U5 s catalytic step
Prp9p, Prp11p		SF3a complex = spliceosome assembly, after U1 mRNA complexing
Prp16p	DExD/H-box	second-step unwindase
Prp17p (Cdc40p)		catalytic step II; cell cycle progression
Prp18p		positioning of 3′-splice site U5 (step II) interact Slu7p
Prp19p		E3 ligase motif
Prp21p		SF3a subcomplex assembly
Prp22p	DExD/H-box	unwinds RNA and lariat intermediates
Prp26p		lariat debranching enzyme
Prp28p	DEAD-box	isomerization of 5′-splice site
Prp38p		conformational changes for catalytic activation
Prp39p		component U1 TPR repeat
Prp40p		U1 second-step branchpoint binding
Prp42p		U1 biogenesis
Prp43p	DExD/H-box	lariat intron release
Prp44p		disruption of U4/U6 base-pairing which activates spliceosome catalysis
Prp45p		interactor for Prp22p and Prp46p
Prp46p		member of NTC (Prp19p) stabilizes U6 RNA in RNP forms of U2, U5, and U6
Sub2p	DECD-box	component of the TREX complex required for nuclear mRNA export; involved in early and late steps of spliceosome assembly

proteins, but lacks an snRNA component. The four C-terminal WD40 segments of Prp19p form a β-propeller structure held together by a central stalk of coiled-coil domains (Ohi *et al.*, 2005). The N-terminus contains a U-box domain that has E3 ubiquitin ligase activity. Homologous genes are conserved across many species, such as human, mouse, *C. elegans*, *D. melanogaster*, *A. thaliana*, and *S. pombe*. Note that in Figure 5.27, the two possible versions of U5 snRNP are distinguished by colors (yellow, with NTC; green, without NTC).

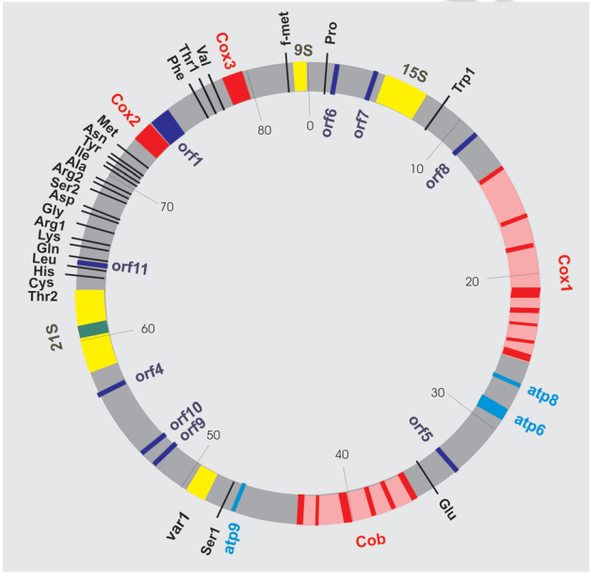


Fig. 5.28 Yeast mitochondrial genome.

Prp2p, an RNA-dependent NTPase and putative RNA helicase, then appears to interact transiently with the spliceosome, activating it for the first transesterification reaction. After ATP hydrolysis, Prp2p dissociates from the spliceosome (King and Beggs, 1990; Kim and Lin, 1996). Upon completion of this first catalytic step, the RNA-dependent NTPase Prp16p joins the spliceosome, interacts with the 3′-splice site, and triggers further rearrangements (Wang and Guthrie, 1998; Schneider *et al.*, 2002). The activities of Prp8p, Slu7p, Prp17p, and Prp18p are also required for completion of the second step (Umen and Guthrie, 1995b; McPheeters *et al.*, 2003; James, Turner, and Schwer, 2002). Two further helicases, Prp22p and Prp43p, are required for release of the spliced mRNA and excised intron, respectively (Martin, Schneider, and Schwer, 2002; Schneider, Campodonico and Schwer, 2004).

Following completion of the splicing reaction, spliceosomes are thought to dissociate and reassemble on other pre-mRNAs for further rounds of splicing. However, pre-assembled complexes have been isolated that, when supplemented with additional factors and pre-mRNA, catalyze splicing, suggesting that spliceosomes need not reassemble *de novo* on pre-mRNA (Stevens *et al.*, 2002; Malca, Shomron, and Ast, 2003). Thus, alternative spliceosome assembly (Maroney, Romfo, and Nilsen, 2000; Nilsen, 2003) or recycling pathways (Verdone *et al.*, 2004) may exist. Splicing efficiency of various transcripts is differentially affected by mutations in

spliceosomal components, such as Prp19p, suggesting that the spliceosome can distinguish between individual transcripts and possibly use these differences to specifically regulate gene expression via control of splicing (Pleiss *et al.*, 2007).

5.4.3 Provenance of Introns

Footprints from an RNA world? Several scientists had speculated that genes originally evolved as exons, and that the progenote organism from which current prokaryotic and eukaryotic organisms evolved may have had a split gene structure (Blake, 1978; Bonen and Doolittle, 1978; Darnell, 1978; Gilbert, 1978). These **primordial exons** are pictured as encoding sequences for stable protein folding domains. Assembly of a number of exon sequences by RNA splicing would be expected to produce a protein composed of stable folding domains that have a high probability of being functional either structurally or catalytically. If genes originally evolved in this fashion, the arrangement of introns in relation to protein secondary structure might not be random. Evidence to support this hypothesis has been sought in the exon–intron structure of evolutionarily old proteins critical for energy metabolism.

Phylogenetic comparison of the sequences of homologous genes from a variety of organisms revealed that intron sequences had drifted much more rapidly than exon sequences. This suggested that intron sequences might generally not be functional, at least in the context of requiring long tracts of specific sequences. Furthermore, the length of introns in homologous genes significantly varied during evolution, suggesting little constraint. Finally, it became clear that specific introns could be lost during evolution. The mechanism responsible for the exact deletion of introns is probably related to gene conversion using a cDNA copy of the mRNA or a partially spliced intermediate RNA. This process has been documented for the removal of introns from yeast genes by Fink (1987) and raises the question of why introns persisted during evolution (Schwartz and Stone, 1990).

However, as discussed in Section 11.2, yeast mitochondria revealed introns in several of its genes completely differing in structure from nuclear genes. One outstanding finding was that introns (at least in part) coded for particular functions, a fact that later also became apparent for many nuclear genes from higher organisms, and another surprise was that mitochondrial introns behaved like mobile genetic elements (Dujon *et al.*, 1986).

5.5 Extrachromosomal Elements

5.5.1 Two Micron DNA

About 40–60 copies of stably maintained, circular 2 μm ("two micron") DNA plasmids are found in the nucleoplasm of most common strains of *S. cerevisiae* cells (overviews: Futcher, 1988; Parent, Fenimore and Bostian, 1985). The plasmid replicates exactly once in each S phase of the cell cycle; the DNA is packed into histones. The plasmid is relatively small (6318 bp). Its presence confers no obvious advantage to its host nor does it appear to impose any disadvantage at its steady-state copy number. The plasmid contains four protein-coding loci (FLP1, REP1, REP2, and RAF1) and four *cis*-acting loci (an origin of replication, a partitioning locus called STB, and two Flp recombination targets (FRTs).

The usefulness of 2 μm DNA in the construction of yeast shuttle vectors has been pointed out in Section 4.2. Plasmids structurally related to 2 μm DNA have been observed in *Zygosaccharomyces* and *Kluyveromyces* species (Bianchi *et al.*, 1987; Wickner, 1995).

5.5.2 Killer Plasmids

Certain killer strains of *S. cerevisiae* harbor double-stranded RNA plasmids as extrachromosomal elements (Wickner, 1996). The killer phenomenon – a well-known yeast–yeast interaction – is due to the fact that these strains secrete proteinaceous toxins that are lethal to sensitive strains, but to which the killer strains themselves are immune. Killer yeasts are widespread among laboratory cultures as well as in natural habitats. Many yeasts other than *S. cerevisiae* exhibit a quite different genetic basis for the killer phenomenon. For example, killer toxins from *Kluyveromyces lactis* or *Pichia acaciae* are encoded by linear DNA plasmids, while in *Pichia kluyveri* or *Hansenia uvarum* they are chromosomally inherited.

5.5.3 Yeast Prions

Early signs of ominous guests? Most importantly, all yeast strains carry three unusual, protein-based genetic elements that have been classified as prions (Tuite and Lindquist, 1996; Uptain and Lindquist, 2002). They alter cellular phenotypes through self-perpetuating changes in protein conformation and are cytoplasmically partitioned from mother cell to daughter. These prions affect diverse biological processes: translational termination, nitrogen regulation, inducibility of other prions, and heterokaryon incompatibility. Two of the yeast prions were described very early, [PSI^+] by B.S. Cox (Cox, 1965) and [URE3] by F. Lacroute (Lacroute, 1971); the third, [RNQ^+], has been identified in S. Lindquist's laboratory (Sondheimer and Lindquist, 2000).

[URE3] is the prion form of the yeast Ure2p protein (Wickner, 1994; Wickner, Masison, and Edskes, 1996) involved in nitrogen metabolism (Masison and Wickner, 1995). The first 65 amino acids (the so-called UPD domain) are dispensable for function, but necessary and sufficient for amyloid fiber formation (Schlumpberger *et al.*, 2000; Schlumpberger *et al.*,

2001). cDNA microarray studies have shown that [URE3] cells fail to elicit a transcriptional response of many genes (Ross and Wickner, 2004).

The yeast prion [PSI$^+$] is generated from a conformationally altered form of Sup35p (Paushkin *et al.*, 1996; Derkatch, Bradley, and Liebman, 1998) – one of the two proteins that constitute the translational release factor, eRF3. Some mutant Sup35 proteins cause ribosomes to read through stop codons at an appreciable frequency (Eaglestone, Cox, and Tuite, 1999). Such mutants suppress nonsense codon mutations in other genes, hence their name. [PSI$^+$] strains also display a nonsense-suppression phenotype because translational termination becomes impaired when Sup35p adopts the prion conformation. Curing cells of [PSI$^+$] alters their survival in different growth conditions and produces a spectrum of phenotypes in different genetic backgrounds; [PSI$^+$] provides a temporary survival advantage under diverse conditions, increasing the likelihood that new traits will become fixed by subsequent genetic change (True and Lindquist, 2000). Phenotypic plasticity and the exposure of hidden genetic variation both affect the survival and evolution of new traits (Derkatsch *et al.*, 1996; True, Berlin, and Lindquist, 2004), but their contributing molecular mechanisms are largely unknown. Interestingly, the yeast prions can interact with each other and affect the appearance of other prions (Derkatch *et al.*, 2001; Bradley *et al.*, 2002).

The third yeast prion was identified by a computer search of yeast genome databases guided by its similarity to regions of Sup35p and Ure2p. Rnq1p (the protein determinant of the prion [RNQ$^+$]) received its name through the fact that its sequence is rich in glutamine (Gln, Q) and asparagine (Asn, N) (Sondheimer and Lindquist, 2000). Genetic, biological, and biochemical analyses proved [RNQ$^+$] to be a prion, despite being caused by a nonessential protein of unknown function. Moreover, [RNQ$^+$] is similar to an epigenetic factor [PIN$^+$] that affects [PSI$^+$] induction, providing evidence that one prion affects the appearance of another. However, *in vivo* [PSI$^+$] and [PIN$^+$] form separate structures in yeast (Bagriantsev and Liebman, 2004). Prion protein gene polymorphisms for *RNQ1* have been observed in strains that carry a particular 19-bp deletion in Sup35p. The expansion and contraction of DNA repeats within the *RNQ1* gene thus may offer an evolutionary mechanism that helps ensure a rapid change between [PRION$^+$] and [prion$^-$] states (Resende *et al.*, 2003).

Some heat-shock proteins, such as the molecular chaperone Hsp104p (Romanova and Chernoff, 2009), members of the Hsp70 family, and their Hsp40 cochaperones (Masison, Kirkland, and Sharma, 2009), are involved in yeast prion propagation. Actually, the *HSP104* gene was isolated in a genetic screen for factors that cure [PSI$^+$] when overproduced (Chernoff *et al.*, 1995). When *HSP104* is highly expressed, most forms of [PSI$^+$] are efficiently cured. Paradoxically, deletion of *HSP104* also cures. Thus, [PSI$^+$] propagation requires an intermediate level of Hsp104p. [URE3] and [RNQ$^+$] are also cured when *HSP104* is deleted, but not when the heat-shock protein is overexpressed.

Hsp104p resolubilizes proteins from aggregates. This activity, which minimally requires Hsp70 and its cochaperone Hsp40, is essential for yeast prion replication. Although it is known how yeast prions can be affected by altering protein chaperones, we lack definite explanations of how these effects may occur mechanistically. Probably, there are many ways in which chaperones interact with each other and with amyloid. In this regard, different paralogs (Ssa1–4p) of the Hsp70 family, have to be considered together with their nucleotide exchange factors Fes1p and Sse1p (Hsp110), cochaperones of the Hsp90 family (Sti1p, Cns1p, or Cpr7p) as well as Ydj1p, Sis1p, and Apj1p as Hsp40 members. In the case of Sup35p, the most efficient factors – obviously directly interacting with the amyloid – have been found to be Ssa1p, and the cofactors Sse1p and Sis1p; *in vitro* the activity is enhanced when they are combined. Hsp104p is also present in the disaggregation complex. Several other examples are discussed in Masison, Kirkland, and Sharma (2009).

More yeast prions can be predicted on the basis of three **genetic criteria**: (i) after curing, the prion can arise again *de novo*, (ii) overexpression of the protein increases the frequency of *de novo* generation of the prion, and (iii) the prion depends on the chromosomal gene encoding the protein for propagation, but the prion has a phenotype similar to that of recessive mutation of the chromosomal genes. Further, the high occurrence of repetitious Gln or Asn residues may assist in prion detection. These repeats, for example, are the only immediately obvious similarity between Sup35p and PrP protein sequences. A recent search in 31 proteomic databases for proteins with at least 30 Gln or Asn residues within an 80-amino-acid region revealed a considerable fraction of eukaryotic candidate proteins. Of the total proteins from *S. cerevisiae*, 107 fell into this category. This estimate was confirmed by another search, obtained with a more stringent algorithm. These putative prions were found involved in diverse biological processes, including transcription and translation factors, nucleoporins, DNA- and RNA-binding proteins, and proteins involved in vesicular trafficking (Uptain and Lindquist, 2002).

Effects of Q/N-rich, polyQ, and nonpolyQ amyloids on the *de novo* formation of the [PSI$^+$] prion in yeast and on the aggregation of Sup35p *in vitro* have been studied (Derkatch *et al.*, 2004) as well as the influence of flanking sequences on polyQ toxicity in yeast (Duennwald *et al.*, 2006). In general, the cellular toxicity of proteins with trinucleotide repeats and their propensity for aggregation and formation of amyloid-like structures have lead to the use of yeast as a convenient model in investigations of neurodegenerative diseases. These aspects are taken up in Chapter 13.

Within the last few years a number of **new prions** have been characterized among the aforementioned putative candidates. [SWI$^+$] is the prion form derived from the chromatin-remodeling factor Swi1p (Du *et al.*, 2008). A prion of yeast metacaspase homolog (Mca1p) was detected

by a genetic screen and designated [MCA^{+}] (Nemecek, Nakayashiki, and Wickner, 2009). Among further proven candidates are two transcription factors – the yeast global transcriptional corepressor protein Cyc8p (Ssn6p) can propagate as a prion, [OCT^{+}] (Patel, Gavin-Smyth, and Liebman, 2009), and the non-Mendelian determinant [ISP^{+}] in yeast is a nuclear-residing prion form of the global transcriptional regulator Sfp1p (Rogoza *et al.*, 2010). A systematic survey identified [MOT3^{+}] as the prion form of the Mot3p protein and illuminated sequence features of prionogenic proteins (Alberti *et al.*, 2009). Among newly identified yeast prions, the prion form of Std3p (involved in the control of glucose-regulated gene expression), called [GAR^{+}], appears to be generated by the protein's interaction with Pma1p (Crow and Li, 2011).

5.6 Yeast Mitochondrial Genome

Cytoplasmic inheritance. Since the discovery by Ephrussi in 1949 of **cytoplasmic heredity** of the respiratory-deficient "petite" mutants, *S. cerevisiae* has been at the heart of mitochondrial genetics. The mitochondrial genes and their mosaic intronic structure were first identified in *S. cerevisiae* and the first mitochondrial gene sequenced was from this organism (*cf.* Chapters 2 and 11).

The multicopy mitochondrial genome from *S. cerevisiae* is characterized by low gene density and high A + T content. Its base composition is highly heterogeneous; while the G + C content of the genes is approximately 30%, the intergenic spacers are composed of quasipure A + T stretches of several hundreds of base pairs, interrupted by more than 150 (G + C)-rich clusters, ranging from 10 to 80 bp in length. These traits explain why scientists have sequenced the genes and neglected the intergenic regions.

The complete sequence of the yeast mitochondrial genome was determined by Foury *et al.* (1998).The genome (Figure 5.28) contains the genes for cytochrome *c* oxidase subunits I, II, and III (*COX1*, *COX2*, and *COX3*), ATP synthase subunits 6, 8, and 9 (*ATP6*, *ATP8*, and *ATP9*), apocytochrome *b* (*CYTB*), a ribosomal protein (*VAR1*), and several intron-related ORFs. The *COX1* and *CYTB* genes contain several introns, some of which are translated, independently or in-frame with their upstream exons, to produce maturases, reverse transcriptases, or site-specific endonucleases (Table 5.11). In addition, the mitochondrial genome contains seven to eight replication origin-like (*ori*), elements and encodes 21S and 15S rRNAs, 24 tRNAs that can recognize all codons, and the 9S RNA component of RNase P. All the genes are transcribed from the same strand, except tRNA^{Thr1}. The mitochondrial DNA sequence of strain FY1679, an isogenic derivative of S288C, is 85 779 bp in length and assembles into a circular contig. Some 10 000 nucleotides are new sequences, essentially composed of long A + T stretches interrupted by many G + C clusters. In agreement with previous estimates, the average G + C content is 17.1%. The *COX1* gene and, to a lesser extent, the *CYTB*, 21S RNA, and 15S RNA genes constitute the largest blocks of higher G + C density. The *ATP6*, *ATP9*, *COX2*, *COX3*, and tRNA genes appear as small G + C-enriched islands in the middle of A + T and G + C cluster-rich regions. The other high G + C density peaks correspond to the G + C clusters, their width depending on the number of these lying near one another.

Table 5.11 Introns in yeast mitochondrial genes.

Gene	Intron group	Mobility
COX1		
ai1	II	yes
ai2	II	yes
ai3	I	yes
ai4	I	yes
ai5α	I	yes
ai5β	I	no
ai5γ	II	no
CYTB		
bi1	II	no
bi2	I	yes
bi3	I	no
bi4	I	no
bi5	I	no
21S RNA		
ω	I	yes

Summary

- This chapter is devoted to a thorough description of the genetic entities and their functions that have been generated by nature to make a small unicellular organism like yeast work as independently and successfully as possible. Some of these structures were recognized early in yeast molecular biology; refinements regarding their functions and interplay became visible only recently.

- At a molecular level, functional sites along yeast chromosomes as well as extrachromosomal elements were characterized. It was observed that the centromeric DNA sequences in all yeast chromosomes – the sites of kinetochore formation and chromosome attachment to mitotic and meiotic spindles – extend over only 200 bp, contrary to the much larger, complex centromeres in *S. pombe* or mammalian cells.

- As early as in 1979, it was detected that particular short chromosomal fragments would lend circular plasmids the power to autonomously replicate. Similar

copies of these elements were found to occur frequently along all yeast chromosomes. More intense studies then led to the identification of short loci (ARSs) that form binding sites for the ORC protein complex. Such "prereplicative" complexes are assembled during the M and G_1 phases, persist during the cell cycle, and serve as "markers" for replication origins during the S phase. "Firing" of an origin is initiated by recruitment of the ATP-dependent protein Cdc6 to the ORC complex, which in turn loads other necessary components for chromosome duplication on to chromatin. We consider events that lead to DNA damage and how such obstacles are eliminated. Here, we concentrate primarily on DNA replication and discuss the cell cycle in Chapter 7.

• As a prerequisite for understanding chromosome structure at the most detailed molecular level, the organization of chromatin structure including the histones is discussed. The past years of research have resulted in the discovery of numerous modifications on histones and the enzymes responsible for their deposition. In turn, modifications of histones invoke a reorganization of chromatin structure effected by numerous remodeling complexes, resulting in the promotion or silencing of gene activity.

• In the 1930s, pioneering work by Muller with flies and by McClintock with maize led to the description of telomeres as structures that protect chromosomes from loss and end-to-end fusions. The first telomeres from yeast were cloned in linear plasmid vectors in 1982 and have since served as an indispensable model. Generally, the ultimate ends of eukaryotic chromosomes are composed of reiterated short (G-rich) sequences that bear similarity among different organisms. Owing to their "open-end" structure, a special set of factors is required for their noncanonical, RNA-templated replication. Telomeric DNA and its affiliated proteins serve two crucial functions – they lend stability to the single chromosomes and their structure prevents telomeres from being confused with damaged DNA by checkpoint activities, whose downstream effectors could promote their fusion or degradation, eventually leading to cell cycle arrest and/or cell death. This cell cycle-regulated degradation has been best demonstrated in *S. cerevisiae*, but it probably also occurs in higher eukaryotes.

• The occurrence of transposable elements in yeast was established in 1979. Since then, five different types of retrotransposons have been identified, all of which bear high similarity to retroviruses; some of them were shown to be propagated via VLPs. In contrast to retroviruses, however, these entities are not infectious. Rather, they attracted much attention because they are associated with DNA rearrangements and could be used as models for host–parasite interactions.

• In further sections, we focus on the structures and properties of the yeast cellular RNAs: tRNAs, rRNAs, and mRNAs. In the mid-1960s, efforts to elucidate the genetic code raised an interest in determining the primary structures of tRNAs – those molecules that had been postulated by Francis Crick to function as adaptors in protein synthesis. In 1960, Monier, Stephenson, and Zamecnik devised an approach to isolate low-molecular-weight RNA from yeast by simple phenol extraction and precipitation of the soluble RNA from the aqueous phase with ethanol, which made this organism a most useful source for further work. Fractionation and subsequent analysis of purified tRNA species was much more tedious and took several years. In the end, the first sequence of a tRNA to be determined in 1965 was that of an alanine-specific tRNA from yeast, followed by the sequences of yeast serine, tyrosine, and phenylalanine tRNAs. In all cases, these sequences could be arranged in a "clover-leaf" structure, with the anticodon triplet exposed in the anticodon loop. Some 10 years later, the three-dimensional structure of yeast phenylalanine tRNA was resolved. This model formed a basis to investigate the interactions of tRNA with its cognate partners – the amino acid tRNA synthetases and nucleotide-modifying enzymes. The newly developed molecular techniques, such as cloning and sequencing DNA, were successfully applied to study the genomic arrangement of yeast tRNA genes and to follow the biogenesis of mature tRNA from their precursors.

• Research in 1977 was highlighted by the detection of introns in mammalian mRNAs, but it came as a similar surprise in the same year that yeast tRNA genes also contain "intervening sequences" that have to be processed out from the transcripts during maturation (a procedure that later was confirmed for many eukaryotic tRNA genes in general). Although only some 20% of the nuclear yeast tRNA genes were later recognized to possess intervening sequences, tRNA precursors could successfully be used to characterize the enzymes involved in the cleavage and ligation reactions.

• Although research on yeast ribosomes and ribosome synthesis started in the early 1970s, the fundamental knowledge to this field was mainly contributed from the studies of prokaryotes, preferably the bacterium *E. coli*, or mammalian cells. Nonetheless, the yeast system provided useful details on eukaryotic ribosomal components, on their maturation, and on the regulation of ribosome biosynthesis.

• A field to which yeast made significant contributions was to unravel the mechanism of splicing of eukaryotic pre-mRNAs. Although only 4–5% of the protein-encoding genes from yeast possess introns, a comparison of the "splice cycle" in yeast and mammals revealed great similarity. This finding again underlined the notion that basic

cellular mechanisms and components have been conserved throughout evolution. It took about 15 years (1984–1998) to work out a detailed picture, but there are still novel features to be detected. The spliceosome was recognized as a particle in which the RNA components (pre-mRNA as the substrate and auxiliary snRNAs) were associated with particular proteins (PRPs) forming stable subcellular RNA–protein complexes during the splicing process. In all, over 100 such proteins were characterized. Of invaluable help in defining the single steps within the cycle were a multitude of PRP mutants from yeast, which were defective in splicing because of their inability to assemble specific subcomplexes.

- In addition to the killer plasmids and the 2 μm plasmid, *S. cerevisiae* harbors several unusual, protein-based genetic elements that have been classified as prions. The first two were detected in 1965 and 1971, but ongoing experiments point to the existence of further such elements in yeast.

- Last, but not least, a final section is devoted to the mitochondrial genome, whose organization brought about some surprises. Yeast mitochondria can be obtained as respiratory-competent entities, permitting a functional dissection of respiration, oxidative phosphorylation, and protein import; details of mitochondrial function are presented in Chapter 11.

Further Reading

Araki, H. (2010) Regulatory mechanism of the initiation step of DNA replication by CDK in budding yeast. *Biochimica et Biophysica Acta*, **1804**, 520–523.

Bochman, M.L. and Schwacha, A. (2009) The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiology and Molecular Biology Reviews*, **73**, 652–683.

Clapier, C.R. and Cairns, B.R. (2009) The biology of chromatin remodeling complexes. *Annual Review of Biochemistry*, **78**, 273–304.

Fatica, A. and Tollervey, D. (2002) Making ribosomes. *Current Opinion in Cell Biology*, **14**, 313–318.

Foury, F. and Kucej, M. (2002) Yeast mitochondrial biogenesis: a model system for humans? *Current Opinion in Chemical Biology*, **6**, 106–111.

Grainger, R.J. and Beggs, J.D. (2005) Prp8 protein: at the heart of the spliceosome. RNA, **11**, 533–557.

Granneman, S. and Baserga, S.J. (2004) Ribosome biogenesis: of knobs and RNA processing. *Experimental Cell Research*, **296**, 43–50.

Henras, A.K., Soudet, J., Gerus, M. *et al.* (2008) The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cellular and Molecular Life Sciences*, **65**, 2334–2359.

Hopper, A.K., Pai, D.A., and Engelke, D.R. (2010) Cellular dynamics of tRNAs and their genes. *FEBS Letters*, **584**, 310–317.

Kolitz, S.E. and Lorsch, J.R. (2010) Eukaryotic initiator tRNA: finely tuned and ready for action. *FEBS Letters*, **584**, 396–404.

Lesage, P. and Todeschini, A.L. (2005) Happy together: the life and times of Ty retrotransposons and their hosts. *Cytogenetic and Genome Research*, **110**, 70–90.

Lue, N.F. (2010) The plasticity of telomere maintenance mechanisms in yeast. *Trends in Biochemical Sciences*, **35**, 8.

Natsume, T. and Tanaka, T.U. (2010) Spatial regulation and organization of DNA replication within the nucleus. *Chromosome Research*, **18**, 7–17.

Nilsen, T.W. (2003) The spliceosome: the most complex macromolecular machine in the cell? *Bioessays*, **25**, 1147–1149.

Pillus, L. (2008) MYSTs mark chromatin for chromosomal functions. *Current Opinion in Cell Biology*, **20**, 326–333.

Remus, D. and Diffley, J.F. (2009) Eukaryotic DNA replication control: lock and load, then fire. *Current Opinion in Cell Biology*, **21**, 771–777.

Rodriguez-Navarro, S. (2009) Insights into SAGA function during gene expression. *EMBO Reports*, **10**, 843–850.

Shorter, J. (2010) Emergence and natural selection of drug-resistant prions. *Molecular BioSystems*, **6**, 1115–1130.

Suganuma, T. and Workman, J.L. (2008) Crosstalk among histone modifications. *Cell*, **135**, 604–607.

Verdone, L., Agricola, E., Caserta, M., and Di Mauro, E. (2006) Histone acetylation in gene regulation. *Briefings in Functional Genomics and Proteomics*, **5**, 209–221.

Weake, V.M. and Workman, J.L. (2008) Histone ubiquitination: triggering gene activity. *Molecular Cell*, **29**, 653–663.

Wickner, R.B., Edskes, H.K., Shewmaker, F., and Nakayashiki, T. (2007) Prions of fungi: inherited structures and biological roles. *Nature Reviews Microbiology*, **5**, 611–618.

Zheng, L. and Shen, B. (2011) Okazaki fragment maturation: nucleases take centre stage. *Journal of Molecular Cell Biology*, **3**, 23–30.