UDC 663.12:575.164:577.214.6
ISSN 1330-9862 *review*

(FTB-1051)

Gene Function and Expression: Four Years of the Post-genomic Era of Yeast

*Horst Feldmann**

Adolf-Butenandt-Institut, Molekularbiologie,
Universität München, Schillerstrasse 44, D-80336 München, Germany

Received: October, 20 2000
Accepted: November 29, 2000

Summary

The yeast *Saccharomyces cerevisae* is a key model organism in which eukaryotic cell architecture and fundamental cellular mechanisms can be successfully investigated. The yeast genome sequence programme has yielded a wealth of information on some 6000 genes, half of which could not be detected by conventional approaches. Remarkably, some 1000 of these novel gene functions have been deciphered in the four years of yeast post-genomic era. Further, novel insights have been gained into gene stucture, gene function, protein-protein interactions, and molecular mechanisms of gene expression. This knowledge that has accumulated by the use of genome-wide micro arrays built from the entire set of yeast genes and by other large-scale approaches, is beginning to merge into useful data libraries. Together with the available literature, these will summarize information on yeast functional genomics, such as genome-wide gene knockout, transcript profiling, microarray datasets, results from systematic two-hybrid screens, drug target discovery, and yeast proteomics. Here again, yeast is at the forefront of providing the opportunity to evaluate the impact of genome sequencing on basic molecular and cell biology investigations of this model organism. This review also summarizes aspects of the molecular mechanisms underlying basal transcription by the RNA polymerases in yeast.

Key words: gene expression, gene function, genomics, proteomics, *Saccharomyces cerevisiae*

Introduction

It is now well established that yeast is an ideal system in which cell architecture and fundamental cellular mechanisms can be successfully investigated. Among all eukaryotic model organisms, *Saccharomyces cerevisiae* combines the advantages of being a unicellular organism which, unlike more complex eukaryotes, is amenable to growth on defined media giving the investigator complete control over environmental parameters. Yeast is tractable to classical genetic techniques, and functions in yeast have been studied in great detail by biochemical and novel genomic approaches (overviews: *1*,*2*). In fact, a large variety of examples provide evidence that substantial cellular functions are highly conserved from yeast to mammals.

It is not surprising, therefore, that yeast had again reached the forefront in experimental molecular biology in taking its place as the first eukaryotic organism of which the entire genome sequence has been made available (*3*). The wealth of information obtained in the yeast genome project (*4*,*5*) turned out to be useful as a reference against which sequences of human, animal or plant genes, and those of a multitude of unicellular organisms under study may be compared. Moreover, the ease of genetic manipulation in yeast opened the possibility to functionally dissect gene products from other eukaryotes in this system.

* Corresponding author; Phone: ++49 89 5996 437; E-mail: horst.feldmann@bio.med.uni-muenchen.de

The Yeast Genome Sequencing Project

The yeast genome sequencing project was started in 1989. Between 1992 and 1996, all 16 yeast chromosomes were completed (*1,2* and references therein) and made public by 16 different teams of the international communitiy of yeast scientists. The yeast genome was published on the web by MIPS (Munich Informatics Center) in a fully organized and annotated form, in April 1996. At present, the yeast genome data are well documented by three complementary databases (*6*). Remarkably, a non-ambiguous and non-redundant nomenclature for open reading frames (ORFs), genes and proteins has been agreed upon. In several aspects, the yeast genome has served as a model system for a systematic genome-wide analysis and has been considered a standard for building data libraries (*7*).

Basic Features of the Yeast Genome

Some basic features of the yeast genome should be briefly recapitulated here to lay the grounds for an understanding of gene function and regulation of gene expression in yeast.

Genomic content

Roughly three quarters of the yeast genome sequence (a total of ~12.8 Mb for all 16 nuclear chromosomes) consist of coding sequence, the rest are intergenic regions, which in many cases have been shown to enharbor signals for replication and predominantly for regulation of gene expression. Some ~3 % of the genomic sequence are occupied by the 274 intact tRNA genes [grouped into 42 families, *(8)*] and 59 snRNAs, ~2 % by 51 intact yeast retrotransposons [the five classes of Ty elements, *(8)*]. Detailed maps for the latter are available (http://www.med.uni-muenchen.de/biochemie/feldmann/ or http://www.mips.biochem.mpg.de/yeast/). Otherwise, yeast is poor in repetitive DNA sequences. Ca. 800 Kb have to be added for the 120 tandem copies of rRNA genes located on chromosome XII, and 86 Kb for the mitochondrial genome (*9*). Currently, it is estimated that some 600 nuclear genes (10 % of the total) contribute to mitochondrial biosyntheses and function.

The principal arrangement of ORFs along the two strands of each of the chromosomes is shown in Fig. 1. In about half of the cases, 'adjacent' genes on opposite strands share a common intergenic region believed to bear most of the upstream activator sequences (UAS) or upstream regulatory sequences (URS). In a number of cases, upstream regulatory signals for a given gene may overlap with the coding sequences of a gene on the opposite strand. Clearly, gene arrangement has implications on gene-specific regulation which will be discussed below.

Genetic and physical maps

Prior to the sequencing project, some 1200 yeast genes had been mapped to the different chromosomes (*10*) which in the beginning helped physical mapping and cloning. By and large, the order of genes in the genetic and physical maps was found to coincide sufficiently (*11*). An example for chromosome II is presented in Fig. 2.

The yeast chromosome ends, as in other eukaryotes, exhibit particular interesting features, both genetically and physically, which made them an excellent sytem to investigate telomere biology. Principal structural features (*12*) are summarized in Fig 3. Subtelomeric regions vary between different yeast strains, they show a high plasticity. Several chromosomes have similar repeats in their subtelomeric regions. Among different wild-type strains (*e.g.* brewer's yeasts), multiple copies of *SUC/MAL/MEL* genes occur (*2,5,13*). Particularly, subtelomeric regions contain multiple copies of the *PAU* genes (*14*) (cf. Table 1), the functions of which are not clearly understood.

The telomeres also display special functional features. Several proteins, like the Sir proteins, Rap1p, and Rim1p play a role in silencing particular genes at these sites; it has now been established how Sir proteins are targeted to their sites of action, representing a general

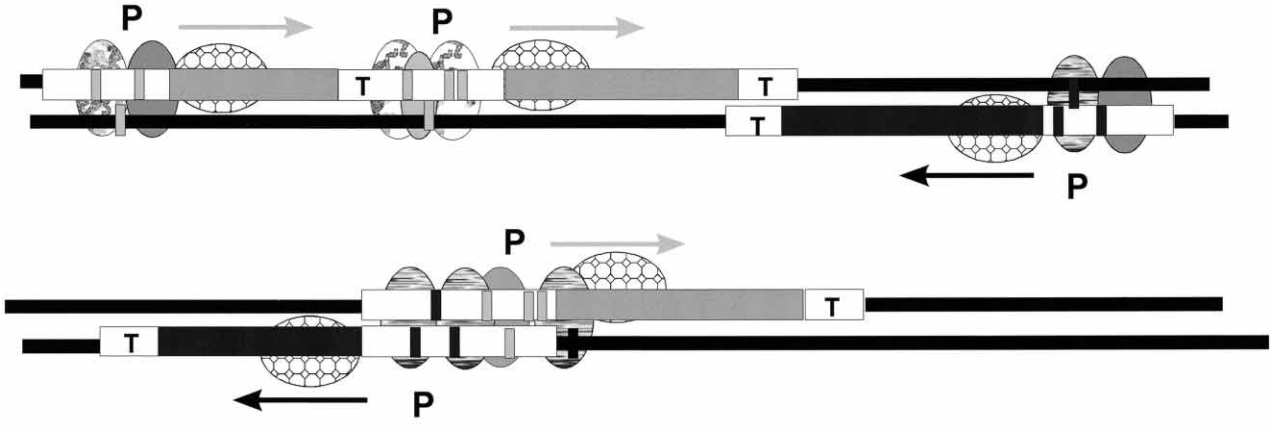

Fig. 1. Principal arrangements of yeast genes and promoters. Coding regions of genes are shown as grey boxes; the open bars represent regulatory regions. Direction of transcription is indicated by arrows. P, promoter; T, terminator signals. RNA Polymerase II (holoenzyme) and gene-specific transcription factors are symbolized by the filled ovals
Key features: Coding sequences: 72 % of each chromosome, Gene density: 1 gene in 2 kb/average. Very few pseudogenes or overlapping genes. (G+C)-content: ORFs = 39.6 %; Intergenic regions = 35.1 %, Lenght of ORFs: <100 aa to >3000 aa; Few genes (~5 %) have introns, Putative membrane spanning proteins: 35 to 40 %, Putative mitochondrial proteins 8–10 %

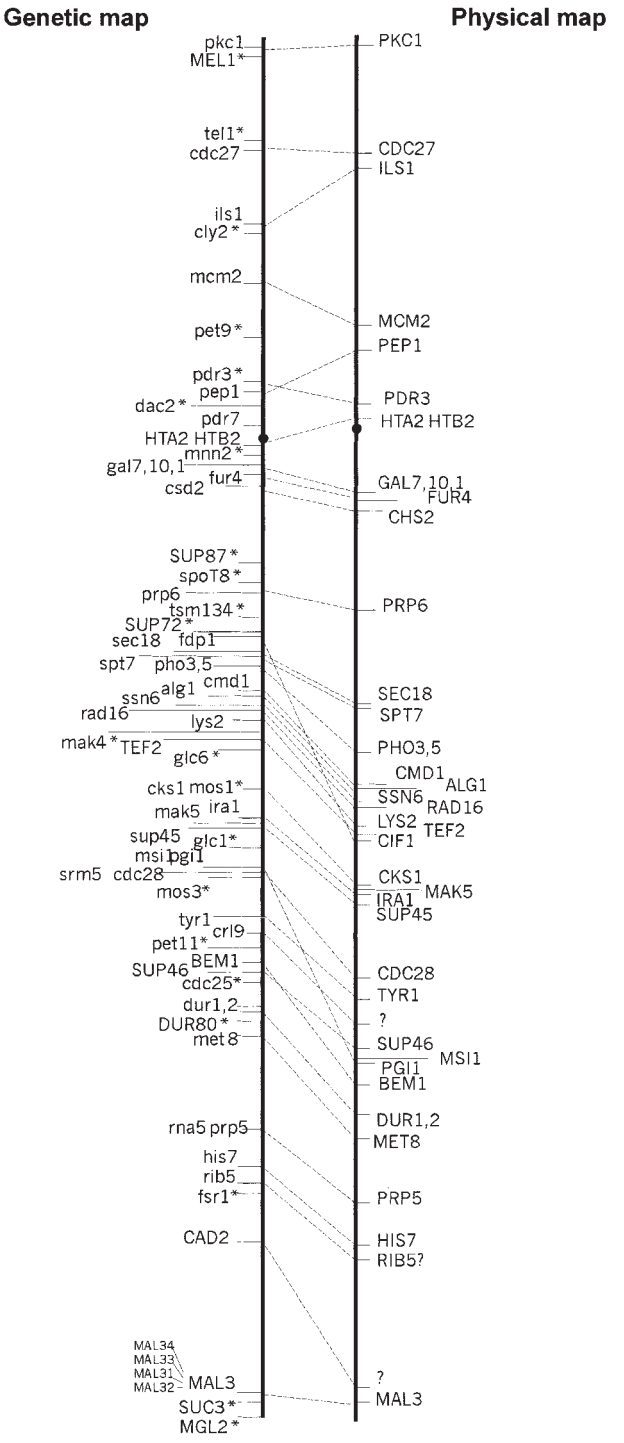

Fig. 2. Genetic and physical maps. Comparison for yeast chromosome II (*13*)

mechanism for regulated repression (*15*,*16*). The telomeres need a special telomerase to be propagated and multiple factors for expression control. There is plenty of literature, but the reader will find ample information in two recent references (*17* and ref. therein).

Coding capacity

The yeast genome comprises some 6000 protein encoding genes. The exact number is still under debate (*1*). A reliable estimate comes from the YPD database (*6*,*18*), which lists 6149 putative proteins, whereof 4270 are characterized by genetics, biochemistry or sequence similarity. MIPS arrives at slightly higher values when including 178 small proteins of less than 100 amino acids (*6*). At present, approximately 32 % of the open reading frames remain of unknown function. This actual value, which has come down from some 50 % of unknown ORFs in 1996, has been reached by intense work on the corresponding genes (see below). Nonetheless, it appears that a certain percentage of an organism's genes are species or genus specific entities, the functions of which can not be disclosed by comparison with those in other organims but only by experimentation. This notion emanates from extended homology analyses among all organisms whose genomes have been sequenced thus far.

Codon selection and tRNA gene content

Interesting observations on codon usage in yeast were made by employing the codon adaptation index (CAI) together with a direct estimate of gene expression (*e.g. 8*,*11*,*19*). CAI is based on determining the amino acid composition and counting the numbers of individual codons for each ORF. Normally, a high CAI value (group I) correlates with a high expression rate for a given gene, whereas a low CAI value (group II) will reflect a low expression rate. All group I genes are known to be highly expressed, whereas low expression levels have not been confirmed for all group II genes. It has been demonstrated that there is a strong correlation between the abundance of yeast tRNAs and the occurrence of the respective codons in protein genes (Fig. 4). In keeping with the general tendency of yeast to avoid GC-rich codons, these are virtually absent from group I genes, while they are used to a similar extent as their synonomous codons in group II genes (*8*).

Gene duplications and genetic redundancy: implications for gene expression

A survey of the sequence data obtained in the yeast sequencing project suggested that there is a considerable degree of internal genetic redundancy in the yeast genome, which on the protein level can be estimated to be around 40 % (*3*,*11*,*20*). Whereas an estimate of sequence similarity (both at the nucleotide and the amino acid level) became predictive at this stage, it still remains difficult to correlate these values to functional redundancy, because even in yeast only a limited number of gene functions have been precisely defined.

In many instances, the duplicated sequences are confined to nearly the entire coding region of these genes and do not extend into the intergenic regions. Thus, the corresponding gene products share high similarity in terms of amino acid sequence or sometimes are even identical and, therefore, may be functionally redundant. However, as inferred from experimental data or suggested by sequence differences within the promoter regions, gene expression should vary according to the nature of the regulatory elements or other (regulatory) constraints (Table 1). It may well be that one gene copy is highly expressed while another one is lowly ex-

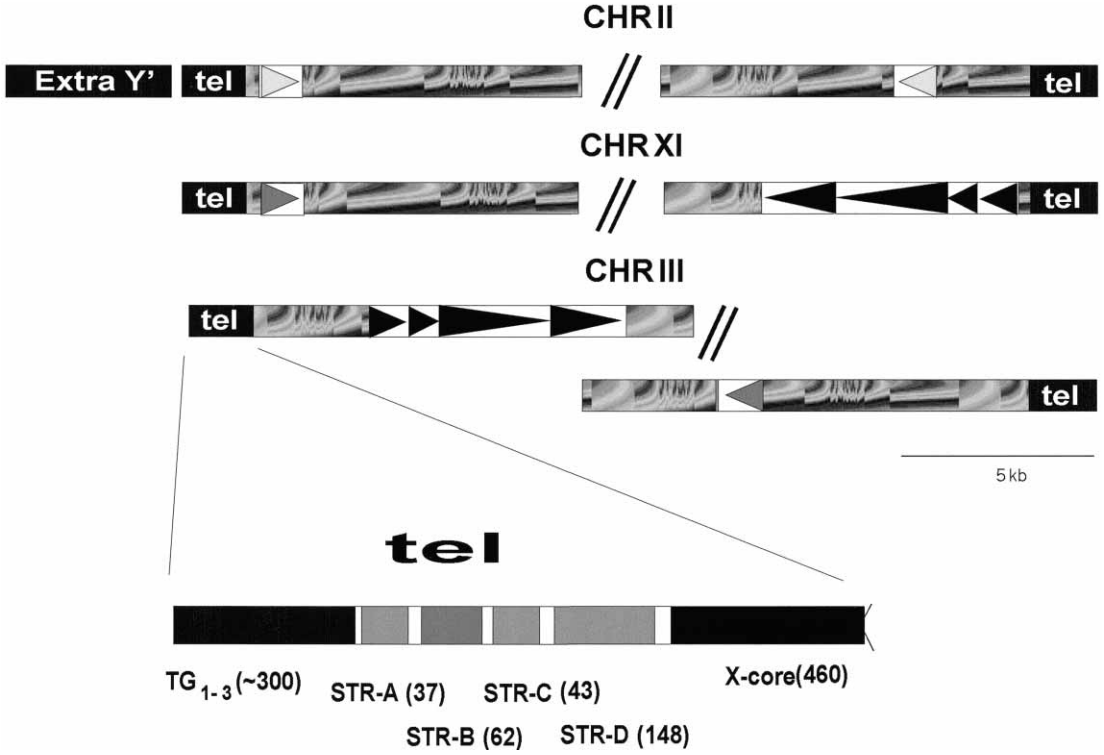

Fig. 3. Telomeres in *S. cerevisiae*. Chromosome ends from chr II, XI, and III are compared. The regular telomere is composed of 300 repeats of TG_{1-3} sequences at the very ends, followed by four similar boxes (STR-A through D) and an 'X-core' of 460 bp. Additionally, some chromosomes carry one of the possible variants of extra Y' sequences (*12*). The arrows indicate inverted repeats, the shaded boxes ORFs with high similarity

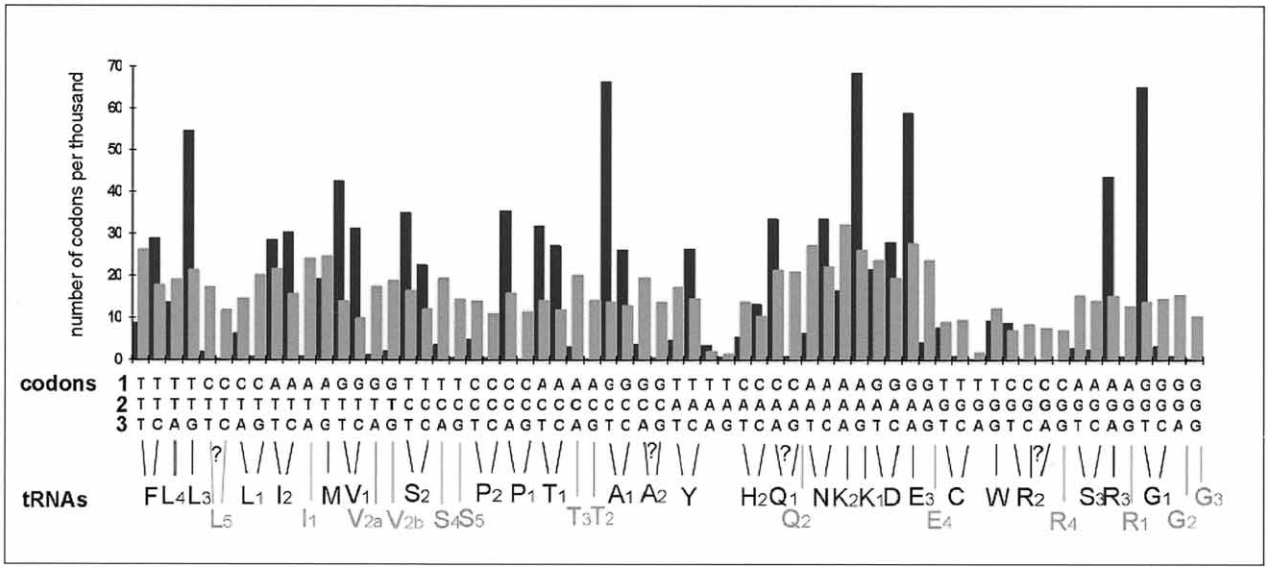

Fig. 4. Codon usage in highly and lowly expressed yeast genes. The tRNAs reading particular codons are identified by the single letter code of the cognate amino acids accepted by them; suffix numbers are used to distinguish isoacceptors (cf. ref. *8*). Black bars, average of 263 highly expressed gens; grey bars, average of 264 lowly expressed genes

pressed. Turning on or off expression of a particular copy within a gene family will depend on the differentiated status of the cell (such as mating type, sporulation, etc.).

Biochemical studies also revealed that in particular cases 'redundant' proteins can substitute each other, thus accounting for the fact that a large portion of single gene disruptions in yeast do not impair growth or cause 'abnormal' phenotypes. This does not imply, however, that these 'redundant' genes were *a priori* dispensible. Rather they may have arisen through the need to help adapt yeast cells to particular environmental conditions.

Table 1. Gene families and gene redundancy

Examples or Type of Genes	Copies in Genome	Protein Similarity	Function	Effects of Multiplicity	Functional Exchange Possible?	Ref.
Major tRNA genes	Up to 12	Identical	Same	Gene dosage	-	8
Minor tRNA genes	1-3	Identical	Same	Gene dosage	-	*8*
Ribosomal proteins	Mostly 2	Identical or very high	Same	Probably	-	*22*
Histones H2A, H2B, H3, H4	2	Identical or very high	Same	Gene dosage	No	*23*
Invertase and maltose metabolism genes (*SUC/MAL*)	Several, strain variation	Identical or very high	Same	Probably gene dosage	Probably yes	*13*
Mitochondrial ADP/ATP carrier (*AAC1/AAC2/AAC3)*	3	High (identical)	Same	Regulation differs	No	*24*
Acid phosphatase *(PHO3/PHO5/PHO10/PHO11)*	5	High (identical)	Same	Regulation differs	No	*25*
Pyruvate carboxylase *(PYC1/PYC2)*	2	Extended	Same	Regulation differs	Yes	*26*
Chitin synthase *(CDS1/CDS2/CDS3)*	3	Extended	Different	Regulation differs	No	*27*
Mannosyl transferase (*KTR* genes)	5	Orthologs	-	Not known	Not known	-
Several 'glycolytic' enzymes	Up to 5	Orthologs	-	Not known	Not known	-
Pleiotropic drug resistance genes (*PDR* family)	Many (11) in family	Orthologs	Similar	All regulated by *PDR1/PDR3*	Probably yes	*28,29*
Amino acid transporters	Several in family	Orthologs	Similar	Regulation differs	Not known	*30*
Sugar transporters	Several in family	Orthologs	Similar	Regulation differs	Now known	*30*
Kinases of various types	Many	Orthologs	-	Now known	Not known	-
Various transcription factors	-	Orthologs	-	Not known	Not known	-
26S Proteasomal *RPT* genes (AAA family)	6	High	Different	Probably concerted control	No (essential)	*31,32*
Mitochondrial ATPase/protease genes (AAA family: *YTA10/YTA11/YTA12*)	3	High	Different	Not down	No	*33,34*
PAU family	Many	High	??	??	Not down	*14*
Metallothionein genes *(CUP1)*	11 tandem	Identical (?)	Same	Gene dosage?	Probably yes	*35*

These notions are of practical importance when carrying out and interpreting gene disruption experiments.

Classical examples of duplicated genes in yeast are the *MEL*, *SUC*, *MGL* and *MAL* genes, which are involved in and have previously been found as subtelomeric repeats in several yeast strains (*21*). In fact, yeast strains differ by the presence or absence of particular sets of these genes. Regarding the genes involved in carbohydrate metabolism, the presence of multiple gene copies could be attributed to selective pressure induced by human domestication, as it appears that they are largely dispensable in laboratory strains (such as αS288C that has served as the common source in the sequencing project) which are no longer used in fermentation processes. Non-homologous recombination processes may account for the duplication of these and other genes residing in subtelomeric regions reflecting the dynamic structure of yeast telomeres in general. Additionally, there is a great variety of genes internal to chromosomes that appear to have arisen from duplications, as suggested by the analyses of individual chromosomes (*3*,*4*).

Remarkably, duplicated genes have also been found in clusters (*7*,*36*). Rather unique cases of gene duplications are represented by the large clustered (tandem) gene family of membrane proteins on chromosome I (*37*), and a large cluster on chromosome VIII near *CUP*1 (*35*). The *CUP*1 gene encoding copper metallothionein, is contained in an extended repeat that also includes an ORF of unknown function. The repeated region has been estimated to span 30 kb in strain αS288C, which could encompass 15 repeats, but the number of repeats varies among yeast strains.

An even more surprising phenomenon became apparent, when the sequences of complete chromosomes were compared to each other, revealing that there are large chromosome segments in which homologous genes are arranged in the same order, with the same relative transcriptional orientations, on two or more chromosomes. Obviously, the genome has continued to evolve since this duplication occurred: genes have been inserted or deleted, Ty elements and introns have been lost and gained between the two sets of sequences (*3*,*38*). The occurrence of 55 such Cluster Homology Regions is now manifest for the yeast genome, which were claimed to result from a duplication of the entire yeast genome (*39*). If optimized for maximum coverage, up to 40 % of the yeast genome is found to be duplicated in clusters, not including Ty elements and subtelomeric regions.

At minimum, the clustering of duplicated genes and the occurrence of extended regions of similarity compel us to consider the idea that entire genomic regions were duplicated, followed by rearrangements. These duplication events would appear to be ancient, because the DNA sequence has clearly diverged outside the coding regions; moreover, such clusters even share a number of tRNA genes both in the same location and orientation.

Additionally, other mechanisms have to be implicated to explain the occurrence of single copies of duplicated genes, preferably those found in the subtelomeric regions or as 'maverick genes' outside the cluster regions. One could imagine, for example, that these would represent processed genes that were inserted into the genome relatively recently, a view which is consistent with the

conservation of sequence only in the coding regions. However, all of these duplications would appear to have been created by integration of full-length complementary DNAs, because none appear to be pseudogenes, and this is unexpected in this model. In addition, some of the homologous gene pairs include introns in both genes, which suggest that at least these genes were not duplicated by this mechanism.

Recent analyses of yeast and 13 species of hemi-ascomycetes arrive at a different view stating that probably no ancient duplication of an ancestral genome has occurred (*40*,*41*). Further, among the *S. cerevisiae* genes of known function, no major duplications of metabolic proteins appear to have occurred, while major expansions are observed for genes encoding membrane proteins, factors involved in protein shaping, and in DNA or RNA wielding (P. Slonimski, pers. communication).

Whatever the relative timescale and mechanisms of duplications, these events followed by mutations affecting functional properties give a chance to result in improved environmental fitness. On the other hand, the high gene density in yeast indicates a strong tendency to maintain a compact genome, therefore compensatory mechanisms must exist to remove non-functional or 'superfluous' gene copies.

An interesting problem intimately related to evolution is the origin of the organisational pattern of genes as is manifest to date. However, thus far no criteria are available, be it structural or functional in nature, that could govern the regional arrangement of particular genes. As it appears there is no 'ordered grouping' of genes along the yeast chromosomes, in which, for example, genes with similar expression profiles or functionally related genes involved in a particular metabolic pathway are closely associated to each other.

Four Years of Post-Genomic Research in Yeast

Genomics and gene function

The achievements of 'four years of post-genomic life with 6000 yeast genes' have recently been reviewed (in ref. *1*). The accumulation of a wealth of novel data and the development of new tools and new programmes result from a fruitful cooperation within the yeast scientific community. It is mentioned (*1*) that during this time nearly 5000 scientists in more than 1000 laboratories have produced ~7000 papers on yeast genes and genomics. 1060 yeast proteins, the functions of which were unknown in 1996 (*18*), have now been deciphered. These resources are publicly available in YPD (*6*).

The latest endeavour was to create a platform for the integration of functional genomics with the scientific literature, named *The BioKnowledge™ Library* (*42*), which together with tracking the progress of yeast research found in the literature will summarize information on yeast functional genomics, such as genome-wide gene knockout, transcript profiling, microarray datasets, results from systematic two-hybrid screens, drug target discovery, and yeast proteomics. Here again, yeast is at the forefront of providing the opportunity to evaluate the impact of genome sequencing on basic molecular and cell biology investigations of this model organism. To extend present and forthcoming knowledge to other (higher) organisms, this database aims at incorporating similar data sets from *C. albicans*, *S. pombe*, *C. elegans*, and mammalian genomes. This finally will lead to a common interpretation of data, finding common protein functions, to explain global regulation, and building interactions into cellular pathways. Nonetheless, it may be useful to briefly outline in tabular form genomic studies as they are presently available.

Analysis of gene functions in international programmes

The efficient homologous recombination capability of yeast became a rationale for the construction of a variety of genome-wide deletion libraries, virtually the first to be available. Predominantly, two approaches were chosen (Table 2): *(i)* systematic deletions by the use of kanamycin insertion/deletion cassettes (*43*) that were employed in the EUROFAN (*44*), the German Networks (*45*) and further European programmes, and *(ii)* a 'bar-coded deletion' method (*49*) developed for the Transatlantic Consortium aiming at systematically deleting all 6000 yeast genes and marking all deletants by 20-mer nucleotide signatures, which allow recognition under selective conditions. Results of these projects can be found in the literature cited in Table 2. Also, collections of the respective strains, cassettes, and plasmids are available (Table 2) and are still of use in ongoing projects to analyze particular genes or sets of genes under various phenotypic conditions.

Similarly, a number of libraries have been generated by random insertional mutagenesis of transposable elements into the yeast genome (Table 2), of which the most comprehensive and useful is the recently constructed set of 7800 mutant strains (Tn3 mini transposon insertions) currently analysed by M. Snyder and his collaborators. Together with the laboratory of P. O. Brown at Stanford, they devised a novel method, the so-called chip-chip technology, which combines the usual microarray technique (see below) with a second array of intergenic regions, allowing to hybridize cross-linked transcription factor genes. Using this method, regulatory signals in upstream regions were studied: 163 Swi4p targets, Mbp1 targets, Ser/Thr kinases and Tyr kinases were characterized. Furthermore, the chip-chip approach has been combined with 'protein chips' that are derived from 10x14 microwell arrays in order to detect specific substrates, *e.g.* modified by kinases (M. Snyder, pers. commun.)

Genome-wide microarrays

During the last four years, the approach of fixing large numbers of genes to microarrays (using either glass slides or, in fewer cases, membranes) for subsequent hybridization with cDNA or cellular mRNA has become most popular and independent on industrial manufacturing (Table 3). Pioneering work in this field was done by the laboratory of P. O. Brown at Stanford, who devised a protocol to make the appropriate tools at rather low cost (*52*; http://emgm.stanford.edu/pbrown/mguide/index.html). Thus it became feasible to

Table 2. Generation and analysis of yeast gene libraries

Project	From	Number	Analysis	Ref.	Source available
Delection mutant libraries					
EUROFAN	All chr sequenced by EU	758/4400	758 ORFs of unknown function. systematically under various conditions. Info to Projects: MIPS	*43,44*	EUROSCARF Frankfurt http://www.uni-frankfurt.de/FB/mikro/euroscarf/index.html
German Genome Project	chr 2, chr	322	322 ORFs of unknown function. systematically under various conditions. Info to Project: MIPS	*44,45*	
'Mass murder' approach	chr 11	129 'tandem delections'	217 ORFs	*46*	
Gif-sur-Yvette	chr3	all??	ORFs systematically under various conditions. Info to Project: MIPS	*47*	
Giessen/St. Louis	chr8	all??	ORFs systematically under various conditions. Info to Project: MIPS	*48*	
Transatlantic Consortium	all chr	6400	Marked by unique 20-mer signatures detectable under selective conditions	*49*	RESEARCH GENETICS http://www.resgen.com http://www-deletion.stanford.edu http://sequence-www.stanford.edu/group/years deletion proj
Transposon insertion libraries					
Ty1	chr V?	limited		*50*	
Mu	genome	limited			
mTn-3	genome	7800	Expression levels by lacZ reporter: 27 different phenotypes Subcellular location by GFP/HA tagging	*51*	http://ygac.med.yale.edu/
Gene Families, Data sets					
EUROFAN: Chromosome Structure					JH Hegemann, Uni Düsseldorf http://websrv.mips.biochem.mpg.de/proj/eurofan 2/n2/index
EUROFAN: Sub-cellular Structure and Organelles					L Grivell, Uni Amsterdam grivell@bio.uva.nl
EUROFAN: The Yeast Transportome					B André, ULB de Bruxelles http://muntjac.mips.biochem.mpg.de/eurofan/ytpd/index.html

accomodate the entire set of yeast genes (6400 in an 3,24 cm^2 array of 80x80 spots) to one chip, which can be simultaneously hybridized with two full complements of differently fluorescence-labeled mRNAs, one that has been derived under 'standard' conditions and used as a reference, while the other one is used to monitor changes in expression profiles under varying biological parameters, such as growth conditions (cell states, media), stress conditions, particular deletants or overexpressants ('master' genes, transcription factors, etc.). Evaluation can be effected by a number of appropriate tools (*52*).

Whereas in the beginning attention concentrated on a few selected conditions (Table 3, *52–60*), a recent study analysed 300 complete expression profiles in which transcript levels of 287 mutants and 13 compound-treated cell cultures had been generated (*61*). The immense speed (almost one transcript level analysed per chip per week or day) has already resulted in the generation of nearly 10^{12} data points, which have to await further interpretation.

Rules that have to be taken into account with this type of approach to the yeast 'transcriptome' are discussed in ref. (*1*). Briefly, not only the preparation of DNA probes and culture conditions, possible cross-hybridizations (see '*gene redundancy*') but also time courses of changes in transcript levels have to be considered; in several cases, therefore, kinetics of transcript levels have been studied (*e.g. 52–54*). In too simple approaches, up to 5000/6000 genes can be found to changing the expression profile of the yeast cell. It is also important to note that only changes (in independently reproduced experiments) monitored in the range of 2–10 fold (or higher) can be considered significant. Further, it has to be borne in mind that expression profiles will not necessarily reflect the outcome of gene expression, as post-transciptional events, translation and post-translational modifications of the final gene products are to follow. In this regard, other approaches, such as Northern blots, monitoring the expression of fusion proteins, proteomic or biochemical methods have to be included in studies of gene expression to reach firm conclusions.

Yeast proteomics and two-hybrid analysis

The term 'proteome' has been coined in 1996 (*66*) as the set of proteins from a given cell, tissue or species. As the definition was initially based on the separation of cellular proteins by two-dimensional gel electrophoresis (overview: *67*), thus far only 2000 yeast proteins can be separated; the rest withstands this approach, being integral membrane proteins, small (<100 aa) proteins, or gene products in too low amounts to be detected (*i.e.* less than one fmole). Another obstacle, namely problems in reproducing protein 2D-patterns, seems to have been

overcome by appropriate standard tools and protocols (*68*). The identification of proteins from single spots of 2D gels by mass-spectrometry (overview: *69*) considerably improved the approach, as it is now possible to identify proteins or peptides from minimal amounts of material. Table 4 lists some of the latest achievements. Of course, fine-tuning of activity and stability of the final products by post-translational modifications, such as modifications of particular amino acid residues by phosphorylation, glycosylation, acetylation, *etc*. have to be considered.

Table 3. Generation and analysis of yeast genes by microarrays

Topic	Ref.	Access to data points of analyses
Exploring the metabolic and genetic control of gene expression on a genomic scale	*52*	http://cmgm.stanford.edu/pbrown/explore/
Yeast microarrays for genome wide parallel genetic and gene expression analysis	53	
The transcriptional program of sporulation in budding yeast	*54*	http://cmgm.Stanford.EDU/pbrown/sporulation/
Characterization of three related glucose repressors and genes they regulate in *Saccharomyces cerevisiae*	*55*	
Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization	56	http://cellcycle-www.stanford.edu/
Dissecting the regulatory circuitry of an eukaryotic genome	*57*	http://web.wi.mit.edu/young/pub/regulation.html
An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD		
Dephosphorylation in *S. cerevisiae*		http://web.wi.mit.edu/young/CTD_phosphatase/
Redundant roles for the TFID and SAGA complexes in global transcription		http://web.wi.mit.edu/young/TFIID_SAGA/
Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast		http://web.wi.mit.edu/young/chromatin/
The yeast A kinases differentially regulate iron uptake and respiratory function		http://web.wi.mit.edu/young/PKA/
Ploidy regulation of gene expression	*58*	http://staffa.wi.mit.edu/fink_public/ploidy/
Systematic changes in gene expression patterns following adaptive evolution in yeast	*59*	http://genome-www.stanford.edu/evolution/
Multidrug resistance	*60*	http://www.biologie.ens.fr/fr/genetiqu/puces/publications/pdr1-3_3-7/index.html
Microarrays from Ecole Normale Superiur, Paris		http://www.biologie.ens.fr/en/genetiqu/puces/microarraysframe.html
ScanAlyze		http://rana.Stanford.EDU/software/
Microarrays preparation		http://www.microarrays.org/
Transcription profiles and 2-hybrid screen data		http://www.proteome.com
Microarrays-based approaches: gene functions and drug targets (INK-JET)	*61,62*	http://www.rii.com/tech/pubs/cell_hughes http://www.rii.com/tech/p7bs/natgen_hughes
Microarrays on membranes		http://www.mips.biochem.mpg.de/proj/yeast/
Filamentous growth in yeast (time course profiles)	*63*	
Comprehensive functional analysis by transcript profiling, 2D gels, MS in proteomics at steady-state levels	*64*	
Chip-chip microarray technology	*65*	http://ygac.med.yale.edu/

Table 4. Proteomics: 2D gel electrophoresis, mass spectrometry and 2-hybrid analysis

Topic	Ref.	Access to data points of analyses
Analysis and identification of 300 yeast proteins (MS and other methods)	*70*	http://www.igbc.u-bordeaux.fr/ypm/
Proteins induced upon oxidative stress (proteome level)	*71*	
cAMP mutants (proteome level)	*72*	
Multidrug resistance mutants (proteome level)	*73*	http://www.biologie.ens.fr/fr/genetiqu/puces/publications/pdr1-3-3-7/index.html
Transcription profiles and 2-hybrid screen data (2000 proteins listed)		http://www.proteome.com
Testing $4 \cdot 10^6$ protein combinations in 2-hybrid analysis (11 % of total)	*77*	
Testing $1.2 \cdot 10^6$ protein combinations in 2-hybrid analysis (3 % of total)	*78*	
Testing $3 \cdot 10^6$ protein fragments (-250 aa) as preys against the full complement of yeast genes as baits in 2-hybrid analysis (2.5 % of total thus far)	*76,79*	

Systematic two-hybrid analysis

The impetus for this ingenious method was the idea of identifying physical protein-protein interactions by coexpressing them in the same cell (*74*); yeast was argued to be the best organism for this approach. Since then, this technique has been widely applied and improvements or variations to the original scheme have been devised (*75*). Application of two-hybrid analysis to a systematic, genome-wide screen would afford the analysis of some 36 x 10^6 combinations among all yeast proteins. This has not become feasible as yet, but recently initial large-scale analyses have been successfully carried out (*76–79*). The reader is referred to Table 4 for a listing and references. Remarkably, thousand interactions have been described in the latest project (*76,79*); a new complex being involved in the nuclear spliceosome, as well as in cytosolic mRNA degradation, has been characterized. Of course, interactions as identified by the two-hybrid approach still have to be confirmed and specified by other methods, which is a considerable task for future.

Molecular Mechanisms in Yeast Gene Transcription and Regulation

Basal transcription machineries

Like other eukaryotes, yeast employs three different DNA-dependent RNA polymerases to effect basal transcription. The single entities, as well as all accessory regulatory factors, have been compiled in a web page (http://www.med.uni-muenchen.de/biochemie/YTFD/index.htm). Each of the polymerases share five (smaller) subunits.

RNA polymerase I

Pol I is exclusively reserved to transcribe ribosomal RNA genes: the holoenzyme is composed of 14 subunits and its activity is regulated by upstream activation factors (*80*). A precusursor rRNA molecule is formed and processed into the mature 28S and 18S rRNAs, and 5.8 sRNA (*81*). In yeast, the 5S RNA genes linked to the rRNA loci, are separately transcribed by Pol III, using specific factors (TFIIIA and TFIIIC that bind to the internal promoter elements, see below).

RNA polymerase II

The regulated transcription of the multitude of yeast protein-encoding genes presents a most challenging problem. While a multitude of complexes of the Pol II holoenzyme is involved in committing a gene for transcription (*82*), the basal activity is provided by RNA polymerase II itself, a complex composed of 12 subunits. Fig. 5 shows a summary of the components participating in the game. The implications of chromatin structure, chromatin remodeling (*83*), and participation of histone acetylase complexes (*84*) and deacetylase complexes (*85*) in the regulated expression of protein-encoding genes have been successively characterized during recent years, whereby the yeast system was greatly used to arrive at paradigms. While for particular aspects specific actions and interactions have been described at the molecular level and have lead to three-dimensional models, however, in great part, our understanding remains two-dimensional.

A compelling 'revolutionary' model describing DNA bending and wrapping around RNA polymerase in transcriptional mechanisms has recently been published by Coulomb and Burton (*86*; http://labcoulombe.usherb.ca/). Briefly, the essence of this thourough discussion for Pol II is as follows. The holoenzyme is suggested to interact with both promoter DNA and the accesory transcription factors (TAF_{IIs}, TFIID or SAGA complex in yeast) for initial promoter recognition. Concomitantly, TAF_{IIs} interact with a number of the basal transcription factors. They also play important roles in promoter selection and are particularly important for recognition of several genes encoding cyclins and further genes responsible for cell cycle progression. Remarkably, a subset of the TAF_{IIs} (probably five) have significant sequence similarity to the histones and may build a structure similar to a core histone octamer within the TFIID complex. Further, it has been shown that TAF_{IIs} make many contacts to the general transcription factors.

Based on the current state of knowledge of TAF_{IIs} and the concept of DNA wrapping, the authors (*86*) have presented a model, in which the TAF_{IIs} are involved in an intermediate step of formation of the Pol II pre-initiation complex. It implicates the following requirements: *(i)* TFIID bound to DNA is a form of modified nucleosome core surrounded by a single left-handed loop of DNA. *(ii)* In many cases, a promoter should consist of the TATA-box binding protein (TBP), TAF_{IIs}, and DNA; RNA Pol II and the general transcription factors (GTFs, such as TFIIA, TFIIB, TFIIE, TFIIF, TFIIH) may recognize both DNA and TFIID in promoter binding. *(iii)* Generally, TAF_{IIs} possess repressive transcriptional functions. *(iv)* Most activators and coactivators function by antirepression mechanisms leading to the release of TAF_{IIs} from promoter DNA. *(v)* Some repressors function by locking the TFIID promoter structure. *(vi)* TFIID must clear off the core promoter to assemble the pre-initiation complex, by an exchange towards Pol II and GTFs.

A thourough analysis of components of the basal transcription machinery that participate in Pol II transcription, and to which extention, has been carried out by lab of R. Young by the use of microarrays (*57*, Table 3).

Gene-specific factors, DNA-binding proteins, and promoters in Pol II transcription

Whilst there exists a vast literature on the regulation of particular yeast genes, promoters and factor requirements in their expression (*87*) only a few cases have been detailed in terms of deciphering the underlying molecular expression mechanisms. Two examples are presented in this issue: regulated expression of the the yeast acid phosphatase (*PHO*) genes (*25*), and the yeast Gal4p activation system (*88*).

Our present ignorance on the majority of yeast promoter structures will change slowly. At best, we can currently offer putative DNA-binding sites for some 70

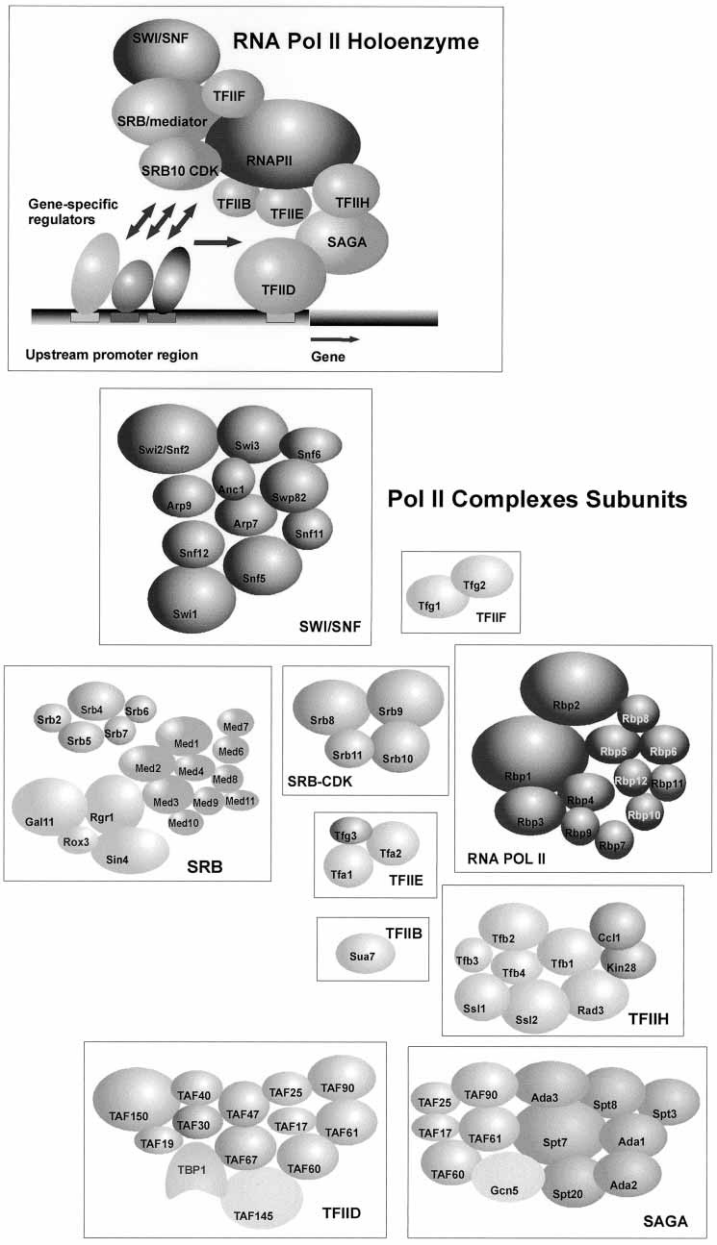

Fig. 5. Polymerase II holoenzyme and subunits of the subcomplexes. Composition of this figure has been modified using information given in ref. *57*

transcription factors, some of which have been determined by experimentation. Knowledge on these items is available but largely buried in the databases. Already along with the yeast sequencing project, we made a compilation of the then known yeast transcription factors, DNA-binding proteins, and the corresponding promoter consensus elements for our consortia (Kleine and Feldmann, 1990, unpublished) and applicable with the GCG Programme (*89*). A similar collection became available in 1995 (*90*). Using information from the various yeast databases, we have now started to develope a more detailed compilation (http://www.med.uni-muenchen.de/biochemie/YTFD/).

Very recently, microarray techniques have started to contribute to the analysis of particular transcription factors and promoter elements (cf. Table 3). In their studies on meiosis, Chu et al. (*54*) have characterized Ndt80p, a protein involved in early induction of meiosis, and Ume6p. Components involved in the yeast pleiotropic drug response have been studied (*60,73*). A thourough analysis of some important yeast factors (Swi4p and Mbp1p) has been reported (*65*) and the data will be available on the web.

RNA polymerase III

Pol III is composed of 14 subunits (*91*). It is responsible for the transcription of the 274 intact tRNA genes, scattered as singular units throughout the genome, and (mostly in concert with Pol II) for most of the 58 snRNAs. The production of the 42 cellular tRNA species occurs through transcription of the corresponding genes

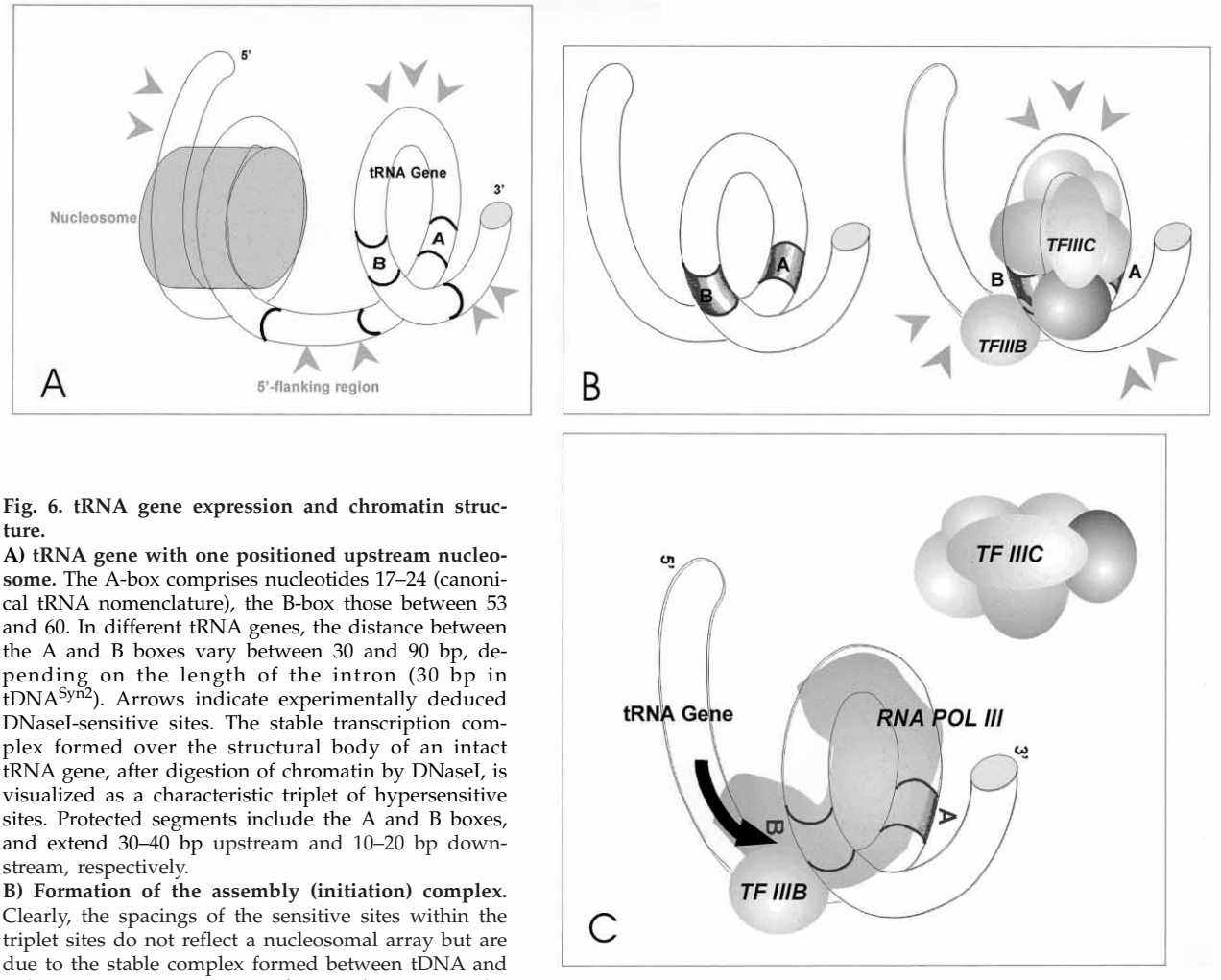

Fig. 6. tRNA gene expression and chromatin structure.
A) tRNA gene with one positioned upstream nucleosome. The A-box comprises nucleotides 17–24 (canonical tRNA nomenclature), the B-box those between 53 and 60. In different tRNA genes, the distance between the A and B boxes vary between 30 and 90 bp, depending on the length of the intron (30 bp in $tDNA^{Syn2}$). Arrows indicate experimentally deduced DNaseI-sensitive sites. The stable transcription complex formed over the structural body of an intact tRNA gene, after digestion of chromatin by DNaseI, is visualized as a characteristic triplet of hypersensitive sites. Protected segments include the A and B boxes, and extend 30–40 bp upstream and 10–20 bp downstream, respectively.
B) Formation of the assembly (initiation) complex. Clearly, the spacings of the sensitive sites within the triplet sites do not reflect a nucleosomal array but are due to the stable complex formed between tDNA and polymerase III transcription factors that protect defined regions from nuclease attack and leave adjacent regions accessible to DNaseI digestion. This interpretation was in good agreement with *in vivo* footprints of the transcription complex and the topographic model (*96*). The first upstream DNaseI-sensitive site that we observed was found located approximately 170 bp upstream from the tRNA gene specific band triplet, and the second upstream site has a similar spacing from this first one. This 'conventional' pattern demands the presence of one or two regularly positioned nucleosomes upstream of the transcriptional complex. Our model was substantiated by analysing $tDNA^{Syn2}$ and mutated versions of it, combined with a variety of different upstream flanking regions. The transcriptional complex formed on intact tDNA flanked by 'favorable' upstream regions was found to act as a strong boundary constraint for the positioning of nucleosomes over the upstream sequences: the first and the second upstream nucleosome reside in defined locations. This array was associated to efficient transcription. Contrary to this, 'non-favorable' sequences found to be unable of forming the regular nucleosomal array resulted in a low transcriptional activity. Upon insertion of a Ty1 delta element into this sequence, however, the array of positioned nucleosomes was completely restored and clearly showed a transcriptional up-modulating effect (*99*). This demonstrates that sequences in the upstream region of the tRNA gene can be changed such that they become amenable for nucleosome positioning. Moreover, efficient transcription of a given tRNA gene appears to depend on sequences far upstream which are not directly involved in the assembly of the transcription complex, but are able to trigger the formation of a transcriptionally suitable chromatin structure. Thus, the sequence-preference in the interactions of histones with the underlying DNA sequences upstream of the tRNA gene is either compatible with an array of positioned nucleosomes or outweighed by the boundary effect. When the A and B boxes in $tDNA^{Syn2}$ were mutated, the transcriptional complex was abolished and the chromatin pattern was changed in a way that the entire region over the modified tRNA gene remained largely free of nucleosomes, particularly no positioned nucleosomes were formed. This may indicate that the upstream sequences in these latter cases exhibit a higher degree of sequence-preference in nucleosome assembly than the above ones, thus competing with the positioning effect induced by the boundary of the transcription complex.
C) Integration of RNA pol III into the initiation complex

and subsequent processing of the precursor molecules at their extended 5' and 3' termini as well as splicing of the precursors derived from the tRNA genes that contain introns (next to the anticodon) (review: *8*). The 5' flanking regions of tRNA genes are preferred target sites for the insertion of Ty elements and multiple transpositions into these 'hot-spots' result in complex patterns derived from sequences of the elements (reviews: *8*,*92*). As a consequence, yeast tRNA genes even those belonging to the same family are embedded in different and rather variable sequence contexts.

RNA polymerase III, unlike RNA polymerase II, is unable to transcribe through a nucleosome (*93*) giving the intragenic placement of the 'assembly' factors (TFIIIC for tRNA genes, and TFIIIA plus TFIIIC for 5S RNA genes) the key role in reserving a gene for transcription. TFIIIB is thought to function as the major anchor in the tDNA complex, marking the initiation site of transcription (*94*). In view of the fact that eukaryotic DNA has to be packaged in a highly condensed nucleosomal complex, a key question related to *in vivo* expression is how the formation of an active transcription complex becomes compatible with chromatin assembly. In this respect and because of their lesser complexity, Pol III transcribed genes have been investigated in the late 80s and early 90s. I will review some of these aspects, as they have not been covered recently.

In tRNA gene transcription, TFIIIC (now established to contain 8 different subunits; *91*) binds to the A and B boxes of the bipartite intragenic promoter (Fig. 6). TFIIIB can then enter the initiation complex. Assembly and topography of the initiation complex have been studied in details (for review see: *95*). In the yeast system, TFIIIB remains tightly bound to DNA during transcription and can direct multiple rounds of transcription, whereas TFIIIC can be stripped off without loss of transcriptional activity. In this regard, TFIIIC takes a role similar to TFIID in Pol II transcription, while TFIIIB is functionally equivalent to the initiation factors of transcription by RNA polymerases I or II, respectively.

There is a variety of evidence that 5'-noncoding sequences can modulate the level of tRNA gene transcription *in vitro* and *in vivo*. *In vitro*, not only promoter strength but also the exact site of initiation have been shown to be influenced by the 5'-flanking sequences (*8*). Ty1 or delta sequences like Ty3 or sigma sequences (*96*) inserted at cognate distances upstream of a tRNA gene were shown to up-modulate its transcription.

To exploit a possible relationsship between *in vivo* transcription and chromatin structure, we developed a system that allows to measure the *in vivo* transcriptional rate of a tRNA gene, which could be connected to different flanking regions, and to analyse the underlying chromatin structure concomitantly (*97*). As a unique reporter for such analyses we built an artificial yeast tRNA gene, $tDNA^{Syn2}$, some 130 bp in length, the body of which was derived from a $tDNA^{Glu3}$ tagged by the insertion of a 'pseudo-intron' of 30 bp that cannot be spliced out from the precursor. The essence of our studies (*97*; Krieg and Feldmann, 1994, unpublished) are summarized in the legend to Fig. 6.

The model we proposed implies that the overall chromatin organisation of the transcription complex correlates with the efficacy of tRNA gene expression: efficient transcription depends on the formation of the stable transcription complex and an array of upstream positioned nucleosomes (Fig. 6A). Clearly, the formation of the transcriptional complex in this model is very similar to the model of DNA wrapping in Pol II transcription (*86*).

Around that time, our findings were corroborated by two reports. Morse *et al.* (*98*) analyzed yeast transformants in which a yeast tRNA gene was fused to strong nucleosome positioning signals such that the predicted nucleosome structure would incorporate the tRNA start site and essential A-box element near to its center. They found that competent tRNA genes were transcribed *in vivo* and were not incorporated into positioned nucleosomes, whereas mutated, inactive tRNA genes were incorporated into nucleosomes whose positions were as predicted. This finding demonstrates that the transcriptional competence of the tRNA gene determined its ability to override a nucleosome positioning signal *in vivo* and establishes that a hierarchy exists between *cis*-acting elements and nucleosome positioning signals. In the modified tRNA gene constructs we used, the flanking sequences obviously did not contain strong nucleosome positioning signals, because the inactive tRNA gene was not incorporated into nucleosomes as inferred from its high accessibility to cleavage by DNaseI.

Burnol *et al.* (*99*) demonstrated that, after nucleosome reconstitution or chromatin asssembly on a $tRNA^{Glu}$ gene, transcription became dependent on TFIIIC and on the integrity of the B-block element. The authors concluded that TFIIIC is related to chromatin repression rather than to transcription complex assembly and seems to be required to help transcription factors to gain access to DNA and to stabilize productive transcription complexes. For example, binding of TFIIIC to the B block could perturb chromatin structure, thus allowing assembly of TFIIIB components at the upstream promoter.

Several explanations could be envisaged to interpret the above results. One possibility is that TFIIIC takes a dual role, namely to clear the tRNA transcriptional unit of a repressive nucleosome structure and, once it has been bound to the internal promoter elements, to facilitate the entry of the other components of the transcription complex. This type of mechanism had been proposed in several instances of polymerase II transcription units (for review, *100*). This view anticipated the analogy of TFIIIC to TFIID.

Second possibility is that a competition occurs between nucleosome positioning and transcription factor binding. The findings of Morse *et al.* (*98*) clearly showed that in the case of an intact tRNA gene factor binding dominates. The reason for the dominance of the cognate transcription factors (TFIIIB and TFIIIC) may lie in their (temporal or spatial) cellular abundance or in a higher affinity for the same DNA sequences as compared to the histones. In any case, transcription of tRNA genes seems to be more resistant to nucleosome-mediated repression than are class II or 5SRNA genes.

From a functional viewpoint, a third possibility can be envisaged to explain the dominance of the tRNA gene transcriptional unit, namely that the assembly of nucleosomes on tDNA is excluded *a priori*. Our findings, that a mutated tRNA gene is largely free of nucleosomes, would favor this assumption. Nucleosome exclusion might be due to the exceptional high (G+C) content of the tRNA genes, which thus might represent sort of nucleosomophobic sequences to particular structural features of tDNA (*e.g.* DNA-bending), or to additional factors involved in complex formation.

Changes in nucleosomal arrays, as observed for the $tDNA^{Syn2}$ variants (*97*), possibly reflect local changes in nucleosome affinities for particular sites, due to alterations of DNA structure. Indeed, »statistical sequencing« of core DNA points to sequence-dependent variation in flexibility as one factor in nucleosome affinity. As these parameters cannot be predicted from a particular sequence, systematic alterations entailing deletions or substitutions of the flanking region of a particular tRNA gene would not have been meaningful. Even subtle modifications could completely change the structural context, and therefore, not only *in vitro* but also *in vivo* studies of this kind have to be viewed with caution. The great advantage of the approach we have taken was that transcription efficiency could be directly determined and correlated with chromatin structure.

The modulatory effects on *in vivo* tDNA expression of particular sequences in our constructs were moderate (no greater than 5-fold). One has to consider, however, that most of the tRNA genes occur in multiple copies, so that gene dosage together with modulation might in effect adapt the number of tRNA transcripts to actual growth requirements. It is also conceivable that the expression of a tRNA gene could be diminished or even completely repressed through particular flanking sequences that contain a strong nucleosomal positioning signal and thus could overcome the positioning by the boundary mechanism, although this might not be very meaningful under natural conditions. Artificially, a region-specific effect has been demonstrated by placing a tRNA gene at HMR. Normally, expression of the genes at HMR is repressed, and so is the expression of the tRNA gene at this location (*98*). Apparently, the repression of the silent mating type loci is mediated by the chromatin structure and, especially, dependent on intact histone H4.

From our experiments it was not possible to discriminate which of the tRNA transcription factors is the particularly active element in positioning upstream nucleosomes. With a view to the antirepression effect of TFIIIC (*99*) and the outcome of our experiments, one could argue that TFIIIC takes the dominant role. Depending on the upstream sequences, an array of nucleosomes is then generated which is more or less favorable for TFIIIB to enter the transcription complex. Only when the nucleosomes become accurately positioned, the complex will be formed and stabilized in a way to guarantee optimal transcription. In the light of more recent findings, one might even anticipate that the regular array then allows for an interaction of the histones with a component of the transcription complex, possibly TFIIIB, and that by this the nucleosomes have an activating role in tRNA gene expression.

5S RNA transcription

For the 5S RNA genes, it has been shown that competition can occur between transcription complex assembly and chromatin assembly on replicating DNA *in vitro*. If nucleosome assembly is predominant, this leads to a selective repression of the 5S RNA genes, which is viewed as an important variable in determining gene activity (*101*). *In vitro*, the formation of transcriptionally competent yeast 5S RNA gene chromatin is dependent on the preformation of a complex consisting of all three transcription factors; if, by contrast, yeast 5S DNA is first assembled into chromatin with core histones, the gene is inaccessible to the polymerase III gene transcription machinery. Chromatin-mediated inhibition of the 5S gene takes place under conditions, in which tRNA genes are transcribed. On the other hand, tRNA gene transcription remains unaffected by partial loss of nucleosomes in yeast under conditions in which several class II genes are transcriptionally activated.

Outlook

The yeast genome sequence programme has yielded a wealth of information on some 6000 genes, half of which would not have been detected by conventional approaches. Remarkably, some 1000 of these novel gene functions have been characterized in the four years of yeast post-genomic era. Further, novel insights have been gained into gene stucture, gene function, protein-protein interactions, and molecular mechanisms of gene expression. This knowledge has largely accumulated by the use of genome-wide microarrays built from the entire set of yeast genes and other large-scale approaches, and is now beginning to merge into useful data libraries. Together with the available literature, these will summarize information on yeast functional genomics, such as genome-wide gene knockout, transcript profiling, microarray datasets, results from systematic two-hybrid screens, drug target discovery, and yeast proteomics. Here again, yeast is at the forefront of providing the opportunity to evaluate the impact of genome sequencing on basic molecular and cell biology investigations of this model organism. In this review, particular attention has been devoted to salient aspects of the molecular mechanisms underlying basal transcription by the RNA polymerases in yeast.

Although we are still far from understanding of how a simple eukaryotic cell might work, we are now on the way to compile information on regulatory circuits and metabolic networks, in a multitude of physiological conditions or environmental constraints. Finally, it will need a high imaginative power to be able to conceive how all these processes and trafficking occur in a cell packed with interacting molecules.

Acknowledgments

Our work has been supported by grants from Deutsche Forschungsgemeinschaft (SFB 190, München), Fonds der Chemischen Industrie, EU BRIDGE and

EUROFAN programmes. I wish to thank all collaborators and colleagues who have contributed to this work over the past years.

References

1. A. Goffeau, *FEBS Lett. 480* (2000) 37–41.
2. H. Feldmann: The yeast *Saccharomyces cerevisiae*: Insights from the First Complete Eukaryotic Genome Sequence. In: *Molecular Fungal Biology*, R. P. Oliver, M. Schweizer (Eds.), Cambridge University Press (1999) pp. 78–134.
3. A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H.-W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S. G. Oliver, *Science, 274* (1996) 546–567.
4. A. Goffeau, R. Aert, M. L. Agostini-Carbone, A. Ahmed, M. Aigle *et al.*, *Nature, 387* (1997) (suppl.) 1–105.
5. H. Feldmann: *ICRF Handbook of Genome Analysis,* N. K. Spurr, B. D. Young, S. P. Bryant (Eds.), Chapter 30, Blackwell Science Publications, Oxford (1997).
6. MIPS (http://www.mips.biochem.mpg.de); SGD (http://genome-www.stanford.edu/Saccharomyces) and YPD (http://www.proteome.com).
7. H.-W. Mewes, K. Albermann, M. Bähr, D. Frishman, A. Gleissner, J. Hani, K. Heumann, K. Kleine, A. Maierl, S. G. Oliver, F. Pfeiffer, A. Zollner, *Nature, 387* (1997) (suppl.) 9–32.
8. J. Hani, H. Feldmann, *Nucleic Acids Res. 26* (1998) 689–696.
9. F. Foury, T. Roganti, N. Lecrenier, B. Purnelle, *FEBS Lett. 440* (1998) 325–331.
10. J. M. Cherry, C. Ball, S. Weng, G. Juvik, R. Schmidt, C. Adler, B. Dunn, S. Dwight, L. Riles, R. K. Mortimer, D. Botstein, *Nature (suppl.), 387* (1997) 67–74.
11. M. Johnston, *Trends Genet. 12* (1996) 242–243.
12. E. J. Louis, R. H. Borts, *Genetics, 139* (1995) 125–136.
13. H. Feldmann, M. Aigle, G. Aljinovic, B. Andre *et al.*, *EMBO J. 13* (1994) 5795–5809.
14. M. Viswanathan, G. Muthukumar, Y. S. Cong, J. Lenard, *Gene, 148* (1994) 149–153.
15. F. Palladino, S. M. Gasser, *Curr. Opin. Cell Biol. 6* (1994) 373–379.
16. M. Cockell, M. Gotta, F. Palladino, S. G. Martin, S. M. Gasser, *Cold-Spring-Harb. Symp. Quant. Biol. 63* (1998) 401–12.
17. Y. Tzfati, T. B. Fulton, J. Roy, E. H. Blackburn, *Science, 288* (2000) 863–867.
    J. Zhou, E. K. Monson, S. Teng, V. P. Schulz, V. A. Zakian, *Science, 289* (2000) 771–774.
18. J. I. Garrells, *Nucleic Acids Res. 24* (1996) 46–49; http://www.proteome.com/
19. R. Percudani, A. Pavesi, S. Ottonello, *J. Mol. Biol. 268* (1997) 322–330.
20. B. Dujon, *Trends Genet. 12* (1996) 263–270.
21. R. Stucka, H. Feldmann, In: *Molecular Genetics of Yeast, a Practical Approach*, J. R. Johnston (Ed.), IRL Press, Oxford (1994), pp. 49–64.
22. R. J. Planta, *Yeast*, *13* (1997) 1505–1518.
23. J. Perez-Martin, *FEMS Microbiol. Rev. 23* (1999) 503–523.
24. D. R. Nelson, C. M. Felix, J. M. Swanson, *J. Mol. Biol. 277* (1998) 285–308.
25. P. D. Gregory, S. Barbarić, W. Hörz, *Food Technol. Biotechnol. 38* (2000) 295–303.
26. R. Stucka, S. Dequin, J. M. Salmon, C. Gancedo, *Mol. Gen. Genet. 229* (1991) 307–315.
27. M. Pammer, P. Briza, A. Ellinger, T. Schuster, R. Stucka, H. Feldmann, M. Breitenbach, *Yeast, 8* (1992) 1089–1099.
28. A. Goffeau, J. Park, I. T. Paulsen, J. L. Jonniaux, T. Dinh, P. Mordant, M. H. Sailer jr., *Yeast, 13* (1997) 43–54.
29. A. Decottignies, A. Goffeau, *Nature Genet. 15* (1997) 137–145.
30. B. André (2000) The Yeast Transportome; http://www.mips.biochem.mpg.de/yeast/
31. R. Schnall, G. Mannhaupt, R. Stucka, R. Tauer, S. Ehnle, C. Schwarzlose, I. Vetter, H. Feldmann, *Yeast, 10* (1994) 1141–1151.
32. H. Feldmann, *Food Technol. Biotechnol. 37* (1999) 9–19.
33. R. Tauer, G. Mannhaupt, R. Schnall, A. Pajic, T. Langer, H. Feldmann, *FEBS Lett. 353* (1994) 197–200.
34. T. Langer, *Trends Biochem. Sci. 25* (2000) 247–251.
35. M. Johnston, S. Andrews, R. Brinkman, J. Cooper *et al.*, *Science, 265* (1994) 2077–2082.
36. K. Heumann, C. Harris, H. W. Mewes: Exhaustive Analysis of Genetic Redundancy in *S. cerevisiae*. In: *Proceedings of the Fourth Int. Conf. Intelligent Systems for Molecular Biology*, AAAI Press, Menlo Park (1996) pp. 98–108.
37. H. Bussey, D. B. Kaback, W. Zhong, D. T. Vo *et al.*, *Proc. Natl. Acad. Sci. USA, 92* (1995) 3809–3813.
38. H. W. Mewes, K. Heumann, *Proc. 6th Ann. Symp. on Combinatorial Pattern Matching, LNCS 937* (1995) 261–285.
39. K. H. Wolfe, D. C. Shields, *Nature, 387* (1997) 708–713.
40. J.–L. Souciet *et al.*, *FEBS Lett.*, in preparation.
41. B. Dujon *et al.*, *FEBS Lett.*, in preparation.
42. YPD^{TM}; http://www.proteome.com
43. A. Wach, A. Brachat, R. Pohlmann, P. Philippsen, *Yeast, 10* (1994) 1793–1808.
44. S. G. Oliver, *Trends Genet. 12* (1996) 241–242.
45. K.-D. Entian, T. Schuster, J. H. Hegemann, D. Becher, H. Feldmann *et al.*, *Mol. Gen. Genet. 262* (1999) 683–702.
46. C. Fairhead, A. Thierry, F. Denis, M. Eck, B. Dujon, *Gene, 223* (1998) 33–46.
47. K. J. Rieger, M. El-Alama, G. Stein, C. Bradshaw, P. P. Slonimski, K. Maundrell, *Yeast, 15* (1999) 513–526.
48. R. Niedenthal, L. Riles, U. Guldener, M. Johnston, *Yeast, 15* (1999) 1775–1796.
49. E. A. Winzeler, D. D. Shoemaker, A. Astromoff, H. Liang *et al.*, *Science, 285* (1999) 901–906.
50. V. Smith, K. N. Chou, D. Lashkari, D. Botstein, P. O. Brown, *Science, 274* (1996) 2069–2074.
51. P. Ross-MacDonald, P. S. R. Coelho, T. Roemer, S. Agarwal *et al.*, *Nature, 402* (1999) 413–418.
52. J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science, 278* (1997) 680–686.
53. D. A. Lashkari, J. L. DeRisi, J. H. McCusker, A. F. Namath, C. Gentile, S. Y. Hwang, P. O. Brown, R. W. Davis, *Proc. Natl. Acad. Sci. USA, 94* (1997) 13057–13062.
54. S. Chu, J. L. DeRisi, M. Eisen, J. Mulholland, D. Botstein, P. O. Brown, I. Herskowitz, *Science, 282* (1998) 699–705.
55. L. L. Lutfiyya, V. R. Iyer, J. DeRisi, M. J. DeVit, P. O. Brown, M. Johnston, *Genetics, 150* (1998) 1377–1391.
56. P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. Eisen, P. O. Brown, D. Botstein, B. Futcher, *Mol. Biol. Cell, 9* (1998) 3273–3297.
57. F. C. Holstege, E. G. Jennings, J. J. Wyrick, T. I. Lee, C. J. Hengartner, M. R. Green, T. R. Golub, E. S. Lander, R. A. Young, *Cell, 95* (1998) 717–728;
    M. S. Kobor, J. Archambault, W. Lester, F. C. Holstege, O. Gileadi, D. B. Jansma, E. G. Jennings, F. Kouyoumdjian, A. R. Davidson, R. A. Young, J. Greenblatt, *Mol. Cell, 4* (1999) 55–62;
    T. I. Lee, H. C. Causton, F. C. Holstege, W. C. Shen, N. Hannett, E. G. Jennings, F. Winston, M. R. Green, R. A. Young, *Nature, 405* (2000) 701–704;

J. J. Wyrick, F. C. Holstege, E. G. Jennings, H. C. Causton, D. Shore, M. Grunstein, E. S. Lander, R. A. Young, *Nature, 402* (1999) 418–421;
L. S. Robertson, H. C. Causton, R. A. Young, G. R. Fink, *Proc. Natl. Acad. Sci. USA, 97* (2000) 5984–5988.

58. T. Galitski, A. J. Saldanha, C. A. Styles, E. S. Lander, G. R. Fink, *Science, 285* (1999) 251–254.
59. T. Ferea, D. Botstein, P. O. Brown, R. F. Rosenzweig, *Proc. Natl. Acad. Sci. USA, 96* (1999) 9721–9726.
60. J. DeRisi, B. van den Hazel, P. Marc, E. Balzi, P. Brown, C. Jacq, A. Goffeau, *FEBS Lett. 470* (2000) 156–160.
61. C. J. Roberts, B. Nelson, M. J. Marton, R. Stoughton, M. R. Meyer, H. A. Bennett, Y. D. He, H. Dai, W. L. Walker, T. R. Hughes, M. Tyers, C. Boone, S. H. Friend, *Science, 287* (2000) 873–880;
62. T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stepaninats, D. D. Shoemaker, D. Gachotte, K. Chakraburtty, J. Simon, M. Bard, S. H. Friend, *Cell, 102* (2000) 109–126.
63. H. D. Madhani, C. A. Styles, G. R. Fink, *Cell, 91* (1997) 673–84.
64. F. Baganz, A. Hayes, R. Farquhar, P. R. Butler, D. C. Gardner, S. G. Oliver, *Yeast, 14* (1998) 1417–1427.
65. M. Snyder *et al.*, in preparation
66. M. R. Wilkins, C. Pasquali, R. D. Appel, K. Ou, O. Golaz, J. C. Sanchez, J. X. Yan, A. A. Gooley, G. Hughes, I. Humphery-Smith, K. L. Williams, D. F. Hochstrasser, *Biotechnology (NY), 14* (1996) 61–65;
K. L. Williams, A. A. Gooley, N. H. Packer, *Today's Life Sci. 8* (1996) 16–21.
67. J. E. Celis, M. Kruhoffer, I. Gromova, C. Frederiksen, M. Ostergaard, T. Thykjaer, P. Gromov, J. Yu, H. Palsdottir, N., Magnusson, T. F. Orntoft, *FEBS Lett. 480* (2000) 2–16.
68. A. Blomberg, L. Blomberg, J. Norbeck, S. J. Fey, P. M. Larsen, M. Larsen, P. Roepstorff, H. Degand, M. Boutry, A. Posch, *Electrophoresis, 16* (1995) 1935–1945.
69. J. S. Andersen, M. Mann, *FEBS Lett. 480* (2000) 25–31.
70. M. Perrot, F. Sagliocco, T. Mini, C. Monribot, U. Schneider, A. Shevchenko, A., M. Mann, P. Jeno, H. Boucherie, *Electrophoresis, 20* (1999) 2280–2298.
71. C. Godon, G. Lagniel, J. Lee, J. M. Buhler, S. Kieffer, M. Perrot, H. Boucherie, M. B. Toledano, J. Labarre, *J. Biol. Chem. 273* (1998) 22480–22489.
72. J. Norbeck, A. Blomberg, *Yeast, 16* (2000) 121–128.
73. A. Navrocki, S. J. Fey, A. Goffeau, P. Roepsdorff, P. M. Larsen, in press
74. S. Fields, O. K. Song, *Nature, 340* (1989) 245–246.
75. M. R. Martzan, S. M. McCraith, F. M. Spinelli, F. M. Torres, S. Fields, E. J. Grayhack, E. M. Phizicky, *Science, 286* (1999) 1153–1155.
76. P. Legrain, L. Selig, *FEBS Lett. 480* (2000) 32–36.
77. T. Ito, K. Tashiro, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K.Yamamoto, S. Kuhara, Y. Sakaki, *Proc. Natl. Acad. Sci. USA, 97* (2000) 1143–1147.
78. P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, J. M. Rothberg, *Nature, 403* (2000) 623–627.
79. M. Fromont-Racine, A. E. Mayes, A. Brunet-Simon, J. C. Rain, A. Colley, I. Dix, L. Decourty, N. Joly, F. Ricard, J. D. Beggs, P. Legrain, *Yeast, 17* (2000) 95–110.
80. M. Oakes, I. Siddiqi, L. Vu, J. Aris, M. Nomura, *Mol. Cell. Biol. 19* (1999) 8559–8569.
81. S. Fath, P. Milkereit, A.V. Podtelejnikov, N. Bischler, P. Schultz, M. Bier, M. Mann, H. Tschochner, *J. Cell. Biol. 149* (2000) 575–590.
82. J. R. Warner, *Trends Biochem. Sci. 24* (1999) 437–440.
83. P. D. Gregory, W. Hörz, *Curr. Opin. Cell. Biol. 10* (1998) 339–345.
84. P. A. Wade, A. P. Wolffe, *Curr. Biol. 7* (1997) R82-84;
D. E. Sterner, S. L. Berger, *Microbiol. Mol. Biol. Rev.* 64 (2000) 435–459.
85. A. A. Carmen, G. R. Griffin, J. R. Calaycay, S. E. Rundlett, Y. Suka, M. Grunstein, *Proc. Natl. Acad. Sci. USA, 96* (1999) 12356–12361.
86. B. Coulomb, Z. F. Burton, *Microbiol. Mol. Biol. Rev. 63* (1999) 457–478.
87. J. Zhu, M. Q. Zhang, *Bioinformatics, 15* (1999) 607–611.
88. K. Breuning, *Food Technol. Biotechnol. 38* (2000) 287–293.
89. J. Devereux, P. Haeberli, O. Smithies, *Nucleic Acids Res.* 12 (1984) 387–395.
90. V. V. Svetlov, T. G. Cooper, *Yeast, 11* (1995) 1439–1484.
91. A. Flores et al., *Proc. Natl. Acad. Sci. USA, 96* (1999) 7815–7820.
92. M. J. Curcio, R. H. Morse, *Trends Genet.* 9 (1996) 436–438.
93. R. H. Morse, *EMBO J. 8* (1989) 2343–2351.
94. B. Bartholomew, G. A. Kassavetis, E. P. Geiduschek, *Mol. Cell. Biol. 11* (1991) 5181–5189.
95. N. Hernandez, *Genes Dev. 7* (1993)1291–1308.
96. J. Kirchner, C. M. Conolly, S. B. Sandmeyer, *Science, 267* (1995) 1488–1491.
97. R. Krieg, R. Stucka, S. Clark, H. Feldmann, *Nucleic Acids Res. 19* (1991) 3849–3855.
98. R. H. Morse, S. Y. Roth, R. Simpson, *Mol. Cell. Biol. 12* (1992) 4015–4025.
99. A.-F. Burnol, F. Margottin, J. Huet, G. Almouzni, M. N. Prioleau, M. Méchali, A. Sentenac, *Nature, 362* (1963) 475–477.
100. J. Svaren, W. Hörz, *Curr. Opin. Genetics Dev. 3* (1993) 219–225.
101. G. Almouzni, M. Méchali, A. P. Wolffe, *EMBO J. 9* (1990) 573–582.

Funkcija i ekspresija gena: četiri godine »poslije-genomskog« razdoblja kvasca

Sažetak

Kvasac *Saccharomyces cerevisiae* ključni je modelni organizam za istraživanje građe eukariotske stanice i osnovnih staničnih mehanizama. Program sekvencioniranja genoma kvasca dao je čitavo bogatstvo podataka o približno 6000 gena, od čega polovina nije mogla biti utvrđena uobičajenim načinima. Važno je istaknuti da je oko 1000 novih funkcija gena razjašnjeno tijekom četiri godine »poslije-genomskog« razdoblja. Osim toga, stečene su nove spoznaje o strukturi gena, njihovoj ulozi, protein-protein interakcijama i molekularnim mehanizmima ekspresije gena. Spoznaje dobivene korištenjem mikropretraživanja čitava kvaščeva genoma (*genome-wide microarrays*), a i drugih sličnih pristupa, počinju se povezivati u korisne zbirke podataka. Zajedno s dostupnom literaturom ti će podaci sažeti informacije o kvaščevoj funkcionalnoj genomici, kao što su oni dobiveni sistematskom razgradnjom gena, određivanjem profila transkripcije u stanici, zatim podaci dobiveni mikropretraživanjima (*microarray datasets*), te rezultati sistematskih pretraživanja s pomoću dvohibridnih sustava, te otkrivanje ciljnih struktura za lijekove i kvaščeve proteomike. Ponovno je potvrđeno da kvasac i ovdje ima čelno mjesto u procjeni utjecaja sekvencioniranja genoma na osnovna molekularno-biološka istraživanja ovog modelnog organizma. U ovom je radu također dan pregled molekularnih mehanizama koji osiguravaju tzv. bazalnu osnovu transkripcije s RNA polimerazama u kvascu.