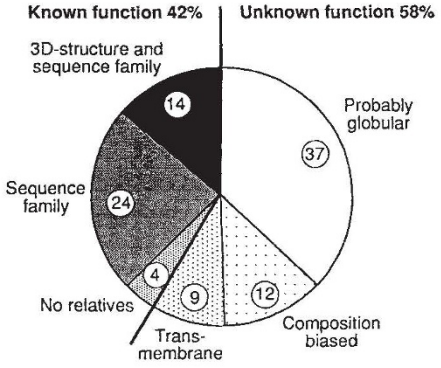
What's in a genome?

SIR — We have taken up the challenge to elucidate the function of the 182 predicted protein products derived from the complete DNA sequence of yeast chromosome III, a result of the European yeast genome project (S. Oliver *et al. Nature* **357**, 38–46; 1992). These authors report that some functional information is available for about 57 of these proteins, either determined by experiment or deduced by similarity searches in sequence databases.

We have identified the probable function of 17 additional protein products, using a combination of low-stringency sequence database searches, various ways of assessing significance, multiple sequence alignment, pattern searches and incorporation of prior knowledge about protein and domain families. The most interesting of these include a DNA polymerase of type X previously found only in mammalia, a new regulatory domain common to eukaryotes and prokaryotes (PILB), a methyltransferase, an acetolactate synthase and a GAL4-type transcriptional activator (see table). In addition, we find that 25 of the chromosome III proteins have homologues of known three-dimensional structure. Taken together, as many as 42% of all proteins of yeast chromosome III have a known or probable function and 13% have an indirectly known three-dimensional structure. Of the remaining 58%, about one-third have putative transmembrane segments (see figure).



Yeast chromosome III proteins. Information accumulated to date by all methods, experimental and theoretical. Information content increases counterclockwise. The principal diversion is between known and unknown biological function. Numbers in per cent. Composition bias indicates unusual amino-acid composition untypical of globular proteins, for example, in coiled coils. The categories are approximate, but give an impression of the current state of the art.

Extrapolating from chromosome III to the entire yeast genome, we can expect that the white, uncharted areas cover only about half of the protein function map but as much as six-sevenths of the protein structure map. As genome projects provide more and more raw sequence data, informatics methods such as those used here can cover much ground, but efficient experimental methods are needed to determine the structure and function of proteins without similarity to known families.

Peer Bork, Christos Ouzounis, Chris Sander, Michael Scharf, Reinhard Schneider & Erik Sonnhammer
European Molecular Biology Laboratory, D–6990 Heidelberg, Germany

SIMILARITY OF SELECTED CHROMOSOME III PRODUCTS TO OTHER PROTEINS

ORF	Length	Family	Closest	%id/len
YCL9c	309	Prokaryotic acetolactate synthases, small subunit	ILVH_ECOLI	36/170
YCL19w	1,347	Transposon B gene family, related to pol genes	COPI_DROME POLX_TOBAC	18/505 20/393
YCL20w	438	Transposon A gene family	YTY1_YEAST	49/439
YCL33c	168	Repressor of pilin promoter	PILB_NEIGO	33/110
YCL75w	146	Pol-like protein	S00954(P)	40/74
YCR14c	582	Type X DNA polymerases	DPOB_RAT	26/393
YCR23c	611	Tetracycline resistance proteins	TCR1_ECOLI	28/150
YVR26c	743	Mammalian PC1 plasma cell membrane protein phosphodiesterase family	PC1_HUMAN PPD1_BOVIN	38/129 38/66
YCR32w	2,167	Hypothetical protein related to C-terminal 'CDC4'-like human fragment	HSCDC4A(E)	49/316
YCR36w	333	Ribokinase (other prokaryotic sugar kinases)	RBSK_ECOLI	38/96
YCR47c	275	ApoMet-methyltransferases	GLMT_RAT	26/301
YCR64c	136	Carboxypeptidases N dipeptidyl-peptidase IV	CBP8_HUMAN DPP_LACLA	27/88 25/105
YCR69w (YCR70w)*	170	Peptidyl-prolyl-*cis-trans* isomerases	CYPH_CANAL	37/122
YCR72c	541	G-protein beta subunits	PR04_YEAST TUP1_YEAST	23/278 32/110
YCR98c	518	Sugar transporter/symporter	A40260(P)	25/179
YCR104w	124	Glucose repressor/cold shock inducible	SRP1_YEAST SCTIPI(E)	27/115 27/99
YCR106w	832	Gal4-like DNA/Zn binding domain	CYP1_YEAST GAL4_YEAST	45/47 19/168

ORF, predicted open reading frame (Oliver *et al.*, 1992); Family, functional protein family; Closest, sequence database identifier of the closest relative(s) from SWISS-PROT (default), PIR (P) or EMBL (E); %id/len, per cent amino-acid identity/length of the alignment.

*When the two adjacent ORFs YCR69w and YCR70w are cut and fused, correcting a possible frameshift sequencing error, they together represent a single member of the proline *cis-trans* isomerase family.

Pyroelectric X-ray generator

SIR — Pyroelectricity is a phenomenon analogous to the more familiar piezoelectricity, a transient voltage produced in response to a change in temperature rather than to a physical stress. Some pyroelectric crystals will also emit electrons[1–7]; electron energies up to 10^5 eV and electric fields as high as 10^6 V cm^{-1} have been obtained in single $LiNbO_3$ crystals. By placing a gold foil close to a pyroelectric crystal of $CsNO_3$, we have made a small X-ray generator which creates photons of about 20 keV.

When $CsNO_3$ is cooled from approximately 300 to 77 K, it produces an electric dipole, which can be maintained at constant temperature because the relaxation time is long[8]. During the cooldown, electrons are ejected from the negative end of the dipole. As the temperature is raised from 77 to about 150 K the dipole begins ejecting electron clusters in bursts from the positive end of the dipole. On warming, electron ejection starts and ends within a specific narrow temperature range for each crystal when the ambient vacuum and rate of change of the temperature is constant.

The sign of the ejected particles was determined to be negative by their deflection in magnetic and electric fields. The particles were detected with a silicon charged particle radiation detector and appeared to have high energy, in the MeV range. However, the beam of particles was attenuated to zero with 1.6 mg cm^{-2} mylar between the crystal and the detector.

When a cluster of electrons with a spread in time less than the resolving time of the system arrives at the detector, the output will reflect the total energy deposited in the detector by the electrons. Therefore, a closely packed cluster of about 15-keV electrons and a 5-MeV α-particle may appear to deposit the same amount of energy in the detector. Using α-emitting sources to calibrate the system, clusters of electrons with an apparent total energy as high as 10 MeV are observed.

In the figure we show the characteristic L X-rays of gold that are produced when a gold foil is adjacent to the positive end of a $CsNO_3$ crystal. A Si(Li) X-ray spectrometer with a 160-eV resolution was used to make this spectrum. These X-rays are produced as the temperature of the crystal is raised from 77 to 300 K and the foil is bombarded with ejected electrons. Fluorescent L X-rays result from vacancies produced in the L shell of Au atoms, which requires an energy of 14.3 keV. However, the K X-rays in Ag, requiring 25.5 keV, were