Direct Allelic Variation Scanning of the Yeast Genome

Elizabeth A. Winzeler,*† Dan R. Richards,† Andrew R. Conway, Alan L. Goldstein, Sue Kalman, Michael J. McCullough, John H. McCusker, David A. Stevens, Lisa Wodicka, David J. Lockhart, Ronald W. Davis

As more genomes are sequenced, the identification and characterization of the causes of heritable variation within a species will be increasingly important. It is demonstrated that allelic variation in any two isolates of a species can be scanned, mapped, and scored directly and efficiently without allele-specific polymerase chain reaction, without creating new strains or constructs, and without knowing the specific nature of the variation. A total of 3714 biallelic markers, spaced about every 3.5 kilobases, were identified by analyzing the patterns obtained when total genomic DNA from two different strains of yeast was hybridized to high-density oligonucleotide arrays. The markers were then used to simultaneously map a multidrug-resistance locus and four other loci with high resolution (11 to 64 kilobases).

Knowledge of genetic variation is important for understanding why some people are more susceptible to disease than others or respond differently to treatments. Variation can also be used to determine which genes contribute to multigenic or quantitative traits such as increased yield or pest resistance in plants or for understanding why some strains of a microbe are exceptionally virulent. Genetic variation can also be used for identification purposes, both in microbiology and forensics, for studies of recombination, and in population genetics (*1*). Rapid and cost-effective ways to analyze variation are needed (*2*).

High-density oligonucleotide arrays have been used to simultaneously measure the expression of every gene in the entire yeast genome (*3*, *4*). These expression arrays contain a total of 157,112 25-mer probes derived from yeast genome coding sequences. Although some regions of the genome have overlapping probes, the arrays cover 21.8% of the nonrepetitive regions of the yeast genome. Because the extent of hybridization of a target sequence to an oligonucleotide probe depends on the number and position of mismatches between the two sequences (*5*, *6*), we hypothesized that a substantial fraction of the allelic variation between any two strains of yeast could be detected simply by hybridizing genomic DNA from the two strains to the arrays and analyzing the hybridization differences (Fig. 1A).

Allelic variation is widespread in different strains and in different individuals in a population. The frequency of variation between common laboratory strains of yeast is estimated to be as high as 1% (*7*). Two *Saccharomyces cerevisiae* strains, S96 (*MATa ho lys5*) and YJM789 (*MATα ho::hisG lys2 cyh*), a clinical isolate from a human lung, were chosen for study (*8*). The strains are phenotypically different—at least five simple genetic loci, including a cycloheximide sensitivity locus from YJM789, can be followed in crosses between these two strains. Partial shotgun sequencing of YJM789 revealed one instance of allelic variation every 160 bases (*9*), with slightly more variation in noncoding regions (*10*). The high degree of array coverage (22%) and the frequency of variation suggested that if only a fraction of the variation could be reproducibly detected, a new genetic map containing a large number of closely spaced markers could be constructed. These markers could then be used to map the loci contributing to the phenotypic differences between the strains.

To test this, we isolated, fragmented, and biotin-labeled genomic DNA from both S96 and YJM789 (*11*). Each sample was hybridized to two different sets of arrays for 2 hours. Then the arrays were washed, stained with a phycoerythrin-streptavidin conjugate, and scanned with a laser confocal scanning device that detects and records the amount of fluorescence at about 3 million physical locations (*3*). Comparison of the images revealed hybridization differences for the two strains (Fig. 1B).

It was anticipated that these hybridization differences could be reproducibly detected and thus could serve as genetic markers. Markers were selected by analyzing the scanned images of arrays hybridized with DNA samples from each parental strain (three times each) and from 14 haploid progeny derived from sporulation of a YJM789/S96 diploid (*12*, *13*). A total of 3714 of the probes on the array were estimated to have greater than 99% probability of being a marker distinguishing the two strains on the basis of their exhibiting a consistent bimodal distribution across all hybridizations. These markers were expected to be from probes whose complementary sequence is completely absent in YJM789 or whose complementary sequence contained a base change near the central region of the oligonucleotide probe. Excluding the ribosomal DNA (rDNA) repeat on chromosome (chr) XII, the average marker spacing was 3510 base pairs (bp). A total of 14 gaps were observed, with the largest gap (59 kb) centered near position 150,400 on chr III (*14*).

To determine whether the set of markers was reliable for linkage analysis, we examined meiotic inheritance. An S96/YJM789 diploid was sporulated, and DNA from four segregants of one tetrad was isolated and hybridized to the arrays. Each of the 3714 markers was assigned a genotype on the basis of whether the observed hybridization signal was closer to the YJM789 or the S96 expected signal response. The probability (p) that the observed signal was of S96 origin was computed (*15*). It was expected that half of the markers would be scored as having an S96 origin and half would be scored as YJM789 and that most markers would segregate with a ratio of 2:2 in the four segregants. The chromosomal locations of the markers, each marker's score (S96 or YJM789), and the location of reciprocal recombination events are shown for one chromosome (XIII) (Fig. 2).

For the entire genome, 97 reciprocal crossovers were observed, close to the expected value of 86 (*16*). For 1220 of the markers, p was less than 0.005 (high probability of YJM789 origin) or greater than 0.995 (high probability of S96) for all four segregants. For this set, 94.5% segregated with a ratio of 2:2; 51% were S96 in origin, and 49% were YJM789 in origin. Some of the markers segregating 3:1 or 4:0 are probably the result of nonreciprocal recombination events, which occur in yeast at frequencies ranging from 0.5 to 30% per locus per tetrad (*17*), consistent with these results. For the remaining markers, p was intermediate (between 0.005 and 0.995) for at least one of the segregants in the tetrad, making it difficult to estimate the frequency of gene conversion.

E. A. Winzeler, D. R. Richards, A. R. Conway, S. Kalman, R. W. Davis, Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305–5307, USA. A. L. Goldstein and J. H. McCusker, Department of Microbiology, 3020, Duke University Medical Center, Durham, NC 27710, USA. M. J. McCullough and D. A. Stevens, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. L. Wodicka and D. J. Lockhart, Affymetrix, 3380 Central Expressway, Santa Clara, CA 95051, USA.

*To whom correspondence should be addressed. E-mail: winzeler@cmgm.stanford.edu

†These authors contributed equally to the work.

Of all the markers (3714), 78.3% segregated with a ratio of 2:2. These data suggest that the probability of misscoring a marker is about 5%, but the probability that a marker will be incorrectly scored for a particular hybridization is strongly correlated with its *p* value and is thus predictable. In studies of single-marker events such as gene conversion or for high-resolution mapping, increased confidence in individual marker accuracy could be obtained by repeating those hybridizations that gave overall low confidence scores (*p*). Even with some noise, a very clear inheritance pattern was discerned, indicating that linkage analysis could be performed with this set of markers.

The YJM789 strain (*MAT*α *lys2 ho:hisG cyh)* and the S96 strain (*MAT***a** lys5 ho) are phenotypically distinguishable. It was predicted that the genomic regions responsible for these differences could be identified by hybridizing DNA from segregants of an S96/YJM789 diploid to the array and analyzing the inheritance of markers. YJM789 (*MAT*α) and S96 (*MAT***a**) are auxotrophic for lysine but have mutations in two different loci: *lys2* (YJM789) and *lys5* (S96) (*18*). YJM789 also carries an insertion in the homothallic mating type locus (*ho:: hisG*) (*19*), whereas S96 has a deletion in the same locus (*ho*). In addition, relative to S96, YJM789 is hypersensitive to multiple drugs, including cycloheximide (*cyh*). The *cyh* locus segregated 2:2 in 99 tetrads of a cross between S96 and YJM789, indicating that a single locus is responsible for the phenotype. Altogether, four known and one unknown loci (*cyh*) could be scored in the cross. The segregants of 99 tetrads were genotyped (*20*). Of the 396 segregants examined, 17 segregants were identified that were *MAT*α *lys2 LYS5 ho cyh*. DNA from 10 of these segregants was hybridized to the arrays and analyzed (*21*, *22*) (Fig. 3).

The most probable parental origin of all DNA segments was determined by estimating the locations of recombination breakpoints for each of the segregants for the entire genome by means of a maximum likelihood method (*23*). This procedure eliminated noise by considering each marker in the context of its neighbors. These data were used to identify regions with a very low probability of random segregation. Probability minima (probability = 0.001 per interval) were located only on chromosomes II, III, IV, VII, and XV (see www.sciencemag.org/feature/data/980398.shl). The physical size of these intervals ranged from 10.7 kb (*LYS2*) to 90 kb (*HO*), with an average genetic size of 17 centimorgans (cM), close to the 20 cM expected (*24*). Four of these regions encompass the known locations of *LYS2* (chr II, 469,702), *MAT* (chr III, 198,278), *LYS5* (chr VII, 215,281), and *HO* (chr IV, 46,272). The *cyh* locus could be unambiguously mapped to the remaining unassigned 57-kb region on chr XV (Fig. 4). These data strongly suggest that *PDR5* (chr XV, 619,838), a multidrug resistance pump (*25*), is the gene responsible for cycloheximide sensitivity. To confirm the

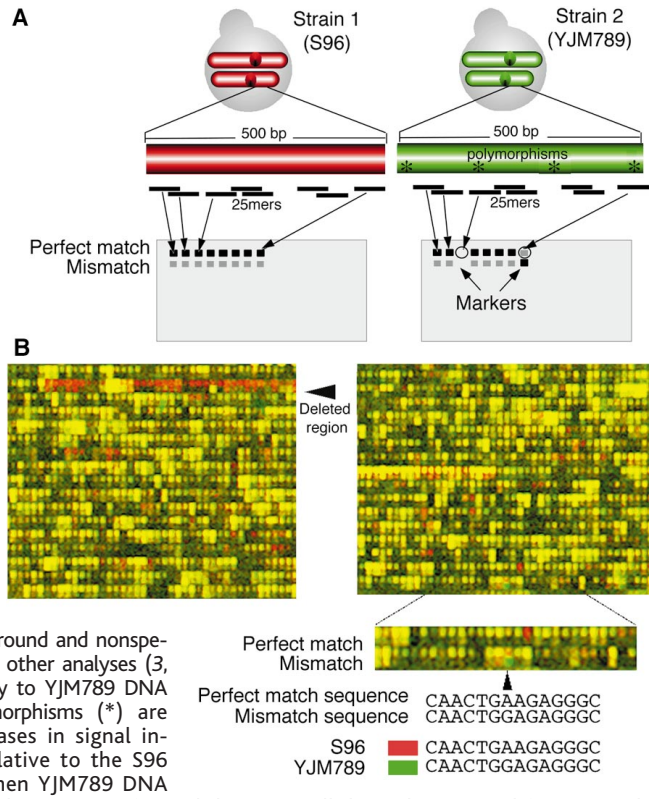

Fig. 1. (**A**) Detecting allelic variation with high-density arrays. For nonduplicated regions of the genome, a minimum of 20 25-base oligonucleotide probes was chosen from yeast genomic sequence (S288c) for every annotated ORF in the yeast genome (*3*). Probes (only from predicted coding regions) were generally arranged on the array in order of their chromosome position. In addition to probes designed to be perfectly complementary to regions of yeast coding sequence (PM), probes containing a single base mismatch in the central position of the oligonucleotide were also synthesized in a physically adjacent position. The mismatch probes serve as background and nonspecific hybridization controls in other analyses (*3*, *31*). If probes complementary to YJM789 DNA fragments containing polymorphisms (*) are found on the array, decreases in signal intensity at these probes relative to the S96 signal may be observed when YJM789 DNA is hybridized to the array. The amount of signal decrease will depend on several factors, such as initial probe intensity and whether the probed fragment is completely absent in YJM789 or contains a small substitution. The location of the polymorphism within the probe sequence will also affect the observed intensity decrease. (**B**) Comparative genomic DNA hybridization patterns. Genomic DNA from two strains of *S. cerevisiae*, YJM789 and S96, was fluorescently labeled and hybridized to two different arrays. Scanned images of the arrays were collected, digitally colored red or green, and then electronically superimposed. A portion of the composite image is shown. Probes that hybridized to S96 DNA more efficiently than YJM789 DNA are red, and probes that hybridize to both DNA samples with equal intensity are yellow. A region that is completely deleted in YJM789 is indicated by an arrow. The figure closeup shows a region in which one of the mismatch features is bright green. Shotgun sequencing of YJM789 demonstrated that the actual sequence of YJM789 was complementary to the sequence of the oligonucleotide in the mismatch row and not to that in the perfect match row.

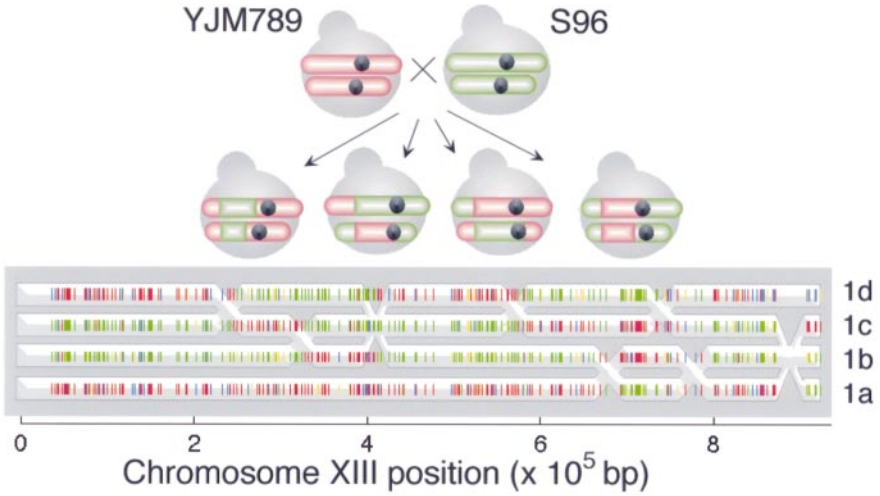

Fig. 2. Inheritance of markers for one chromosome in one tetrad from a cross between YJM789 and S96. Red ticks indicate the location of markers that have a less than 0.5% probability (*p*) of having an S96 origin; blue, $p = 0.5$ to 50%; yellow, $p = 50$ to 99.5%; and green, $p > 99.5\%$.

role of *PDR5* in cycloheximide sensitivity, we deleted the *PDR5* gene in the S96 genetic background and crossed the resulting strain to YJM789. The deletion strain was unable to complement the cycloheximide sensitivity of YJM789 (*26*).

The set of 3714 markers constitutes about 4.7% of the estimated variation between the strains. At 1.0-cM resolution, the map marker density exceeds that of the traditional yeast genetic map (2600 markers) assembled over a period of 40 years (*16*). The high marker density and the fact that all markers can be scored simultaneously should allow the mapping of quantitative or multigenic trait loci (*27*). This method also offers a substantial advantage over any method for scanning or scoring markers described to date: The method does not depend on having probes to the second allele on the array, and because of the sensitivity of the arrays, all markers can be scored in parallel, in a few hours, without amplification steps, gels, or enzymatic manipulation (*6*, *28*–*30*). The method is powerful because of the ease with which genetic markers are identified: A new set of informative markers can be quickly selected for any pair of strains, thus allowing efficient access to the unlimited genetic diversity in the natural world.

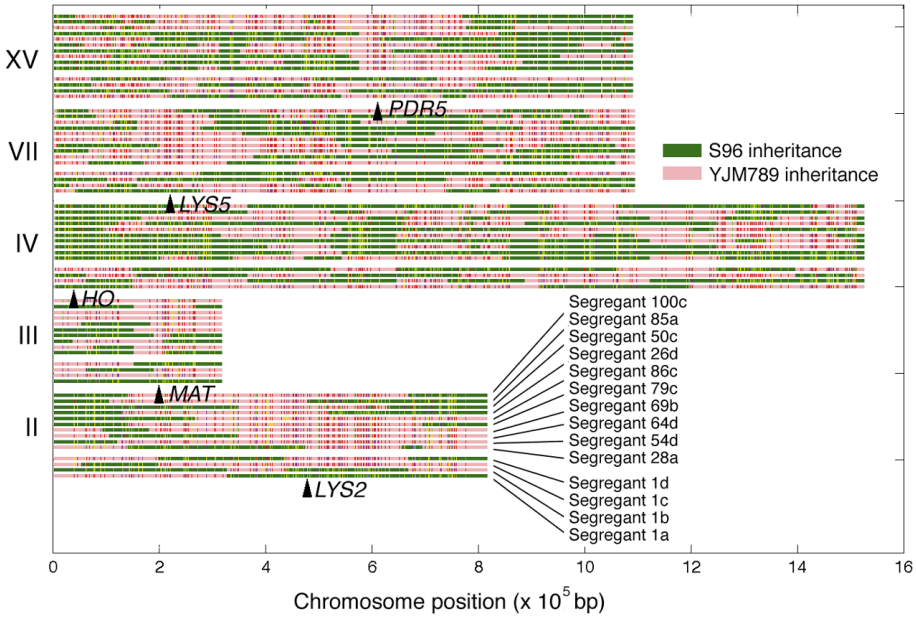

Fig. 3. Inheritance of DNA in 10 segregants for 5 of the 16 chromosomes. Tick colors are as described for Fig. 2. The data are superimposed on a diagram showing the probable location of chromosomal breakpoints, calculated as described in the text. Arrows indicate the known locations of genes. *lys2*, *LYS5*, *MAT*α, and *pdr5* were all inherited from YJM789 (pink), whereas *ho* was inherited from S96 (dark green). All segregants except 1a (*ho pdr5 lys5 MAT***a**), 1b (*ho::hisG MAT*α), 1c (*ho lys2 pdr5 lys5 MAT*α), and 1d (*ho::hisG lys2 MAT***a**) are *ho lys2 pdr5 MAT*α. Data for the entire genome can be found at www.sciencemag.org/feature/data/980398.shl

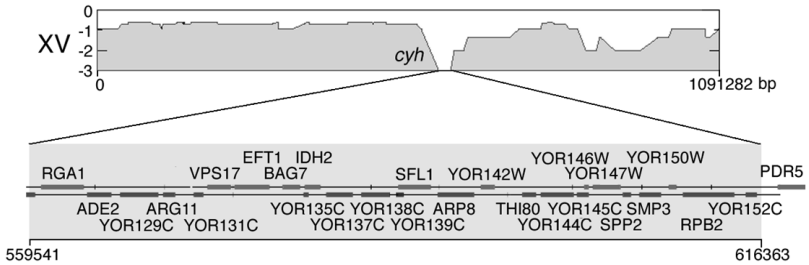

Fig. 4. Calculated probability of random segregation for chr XV. The *y* axis (log base 10) indicates the probability of random segregation calculated with a binomial distribution. The names and locations of ORFs [taken from SGD (*16*)] inside the intervals with the lowest probability of random segregation [10 out of 10 = $(1/2)^{10}$] are shown and are shaded in gray. The minimum interval (559,541 to 616,363) is located just upstream of the *PDR5* gene (619,838 to 624,373) because of a chromosomal breakpoint being assigned to a position 3 kb upstream of *PDR5* for one segregant (86c). Although several markers both upstream and downstream of *PDR5* show S96 inheritance for this segregant, markers from *PDR5* itself were of the YJM789 pattern. The misassignment of the chromosome breakpoint is most likely due to a gene conversion event near the breakpoint. Data for the other chromosomes can be found at www.sciencemag.org/feature/data/980398.shl

References and Notes

1. A. J. Schafer and J. R. Hawkins, *Nature Biotechnol.* **16**, 33 (1998).
2. F. S. Collins, M. S. Guyer, A. Charkravarti, *Science* **278**, 1580 (1997).
3. L. Wodicka, H. Dong, M. Mittmann, M.-H. Ho, D. J. Lockhart, *Nature Biotechnol.* **15**, 1359 (1997).
4. The probes for the entire yeast genome are synthesized in a spatially addressable fashion with a combination of photolithography and solid-phase chemistry [A. C. Pease *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5022 (1994); S. P. A. Fodor *et al.*, *Science* **251**, 767 (1991)], on a series of five 1.64-cm^2 arrays. Each expression array contains more than 65,000 synthesis features, with each physical feature containing more than 10^7 copies of the specific oligonucleotide probe covalently attached to a glass surface. Excluding the rDNA and *CUP1* repeats, the largest gap is 41,325 bases wide at position 510,000 on chr XII.
5. B. J. Conner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 278 (1983).
6. M. Chee *et al.*, *Science* **274**, 610 (1996).
7. S. F. Nelson *et al.*, *Nature Genet.* **4**, 11 (1993).
8. The completed *S. cerevisiae* genome sequence is from strain S288c, and 88% of the S288c genome is derived from EM93, which was isolated from a rotting fig near Merced, California, in 1938 [R. K. Mortimer and J. R. Johnston, *Genetics* **113**, 35 (1986)]. S96 is isogenic with S288c but is unable to undergo mating type switching (*ho*), is able to mate with YJM789, and contains a lesion in the *lys5* gene that can be easily scored in crosses. YJM789 is isogenic with YJM145, a segregant of a clinical isolate of *S. cerevisiae* (*27*). YJM145 has been characterized genetically, and the ultimate source of its parent (human lung) differs substantially from that of S288c in that the strains were isolated from different environments, at different times, and in different geographic locations. Theoretically, any two yeast strains could be used.
9. A library of YJM789 genomic DNA was constructed in an M13 sequencing vector. The sequence was determined for 696 clones as previously described [F. S. Dietrich *et al.*, *Nature* **387**, 78 (1997)]. The sequences were called by means of the phred base caller software (see chimera.biotech.washington.edu/UWGC/tools/phred.htm), which produces a quality measurement for each base [$-10 \times \log(10)$ (probability of an error)]. A total of 122,258 bases were sequenced with greater than 99% confidence by this quality measurement. The YJM789 sequences were compared with S288c sequence with the cross_match program (see chimera.biotech.washington.edu/UWGC/tools/phrap.htm). Discrepancies between the YJM789 and S288c sequences were then classified by quality and assigned into coding and noncoding regions with the phred base caller. In most cases, because only a single trace was available and no alignments were performed, regions of the traces that did not show high quality were excluded from the analysis. When a high-quality sequence (>99.7% accuracy) was used, 466 cases of allelic variation were observed with a frequency of about one every 160 bases.
10. Of the 466 cases of allelic variation in sequences with greater than 99.7% accuracy, 288 were from coding regions (61%). Of the estimated 13.2 Mb, 8.637 Mb (65%) of the yeast genome is annotated as coding sequence by *Saccharomyces* Genome Database.
11. Yeast genomic DNA (10 μg, purified on a Qiagen column) was digested with 0.15 U of deoxyribonuclease I (DNase I) [Gibco-BRL polymerase chain reaction (PCR) grade] in 1× One-Phor-All buffer (Pharmacia) containing 1.5 mM $CoCl_2$ for 5 min at 37°C. After heat inactivation of the DNase I, the DNA fragments were end-labeled in the same buffer by the addition of 25 U of terminal transferase (Boehringer Mannheim) and 1 nmol biotin-N6–dideoxyadenosine triphosphate (NEN) for 1 hour at 37°C. The entire sample was hybridized to the array in a 200-μl volume as previously described (*3*).
12. Grids were aligned to the scanned images by the known feature dimensions of the array. The hybridization intensities for each of the elements in the grid were determined by the seventy-fifth percentile method in the Affymetrix GeneChip software package.

13. An adjusted array hybridization intensity value (*I*) was determined for each hybridization (20 altogether) as the mean of the log(PM) signals of all features that showed minimal variation across all hybridizations (the nonmarkers, determined recursively as described below). Then, for each feature on the array, a linear regression of log[perfect match (PM)] on *I* for all hybridizations was determined by the least squares method, first under the null hypothesis that the S96 and YJM789 samples had the same response and then under the alternative hypothesis that the S96 samples had a greater signal than the YJM789 samples. The models were compared with the *F* test, and the same signal model was rejected in favor of a marker with 99% confidence. This software is available upon request to D. Richards.
14. Gaps were often found near regions with low probe coverage, for example, near repeated elements in the genome or regions of low open reading frame (ORF) density. However, in some cases, probe coverage was adequate, suggesting that the gap might be due to a high amount of sequence conservation or to the region having a recent common origin for the two strains.
15. The *p* is computed as $P(\mathrm{S96})/[P(\mathrm{S96})+P(\mathrm{YJM789})]$, where $P(X)$ is the probability (from the *t* distribution) that a marker has genotype *X*, based on the observed (PM) hybridization signal of the feature and the expected signal (given the array hybridization intensity) and the estimated variance from the regression for the marker.
16. J. M. Cherry *et al.*, *Nature* **387**, 67 (1997).
17. T. Petes, R. Malone, L. Symington, in *The Molecular and Cellular Biology of the Yeast Saccharomyces*, J. Broach, J. Pringle, E. Jones, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1991), vol. I, pp. 407–521.
18. B. Chattoo *et al.*, *Genetics* **93**, 51 (1979).
19. E. Alani, L. Cao, N. Kleckner, *ibid.* **116**, 541 (1987).
20. Yeast strains were routinely grown in yeast extract, peptone, and dextrose (YEPD) medium; sporulation medium and defined medium for scoring auxotrophs were prepared as previously described [F. Sherman, G. Fink, C. Lawrence, *Methods in Yeast Genetics: Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1974)]. Segregants were complementation tested to distinguish *lys2* from *lys5*. Cycloheximide sensitivity was scored by inability to grow on YEPD plates containing cycloheximide (0.5 μg/ml). All segregants from the 99 tetrads were phenotyped for *lys2*, *lys5*, and *cyh*. *lys5* and *cyh* segregated 2:2 for 99 tetrads, and *lys2* segregated 2:2 for 98 of the 99. Selected segregants were scored for *MAT* and *ho*. The *ho* and *ho::hisG* were distinguished by checking the size of PCR products on a gel, and *MAT* was determined by mating and complementation.
21. The loci could have been mapped with any segregant as long as the genotype was known; however, segregants with similar genotypes were chosen to simplify the analysis.
22. The probability of an interval segregating 10 to 0 randomly (a false positive) was estimated to be about 40% for each outcome. No false positives were observed with 10 segregants, and therefore no additional hybridizations were performed. This conservative estimate of probability, which does not take into account recombination hotspots or interference, was calculated by dividing the genome size (12 Mb) by the average interval (29 kb for 10 segregants with 1 cM = 2.9 kb for yeast) and then multiplying this number by the probability of 10 events having the same outcome $(1/2)^N$. In general, up to 13 segregants (or more if the trait is non-Mendelian) may need to be examined to have a 95% probability of identifying a single region as responsible for a trait.
23. The breakpoints were recursively added to each chromosome on the basis of the *p* values. The probabilities of breakpoints at every pair of markers were tested against the probability of no breakpoint. The breakpoints that maximized this likelihood were accepted if the logarithmic likelihood ratio was greater than 30. This procedure was repeated for each new subinterval created by a breakpoint to 500-bp resolution.
24. M. Boehnke, *Am. J. Hum. Genet.* **55**, 379 (1994).
25. E. Balzi, M. Wang, S. Leterme, L. Van Dyck, A. Goffeau, *J. Biol. Chem.* **269**, 2206 (1994).
26. E. A. Winzeler *et al.*, unpublished data.
27. J. H. McCusker, K. V. Clemons, D. A. Stevens, R. W. Davis, *Genetics* **136**, 1261 (1994).
28. D. G. Wang *et al.*, *Science* **280**, 1077 (1998).
29. P. A. Underhill *et al.*, *Genome Res.* **7**, 996 (1997).
30. M. Orita, H. Iwahana, H. Kanazawa, K. Hayashi, T. Sekiya, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2766 (1989).
31. D. J. Lockhart *et al.*, *Nature Biotechnol.* **14**, 1675 (1996).
32. We thank N. Risch and D. Siegmund for helpful advice. E.A.W is supported by the John Wasmuth Fellowship in Genomic Analysis (HG00185-01). D.R.R is a Howard Hughes Medical Institute predoctoral fellow. M.J.M. is supported by a Commonwealth AIDS Research Grants Committee of Australia postdoctoral overseas fellowship. Funding by NIH grant 1R01 HG01633.

28 January 1998; accepted 21 July 1998

Prototype of a Heme Chaperone Essential for Cytochrome c Maturation

Henk Schulz, Hauke Hennecke, Linda Thöny-Meyer*

Heme, the iron-containing cofactor essential for the activity of many enzymes, is incorporated into its target proteins by unknown mechanisms. Here, an *Escherichia coli* hemoprotein, CcmE, was shown to bind heme in the bacterial periplasm by way of a single covalent bond to a histidine. The heme was then released and delivered to apocytochrome c. Thus, CcmE can be viewed as a heme chaperone guiding heme to its appropriate biological partner and preventing illegitimate complex formation.

In c-type cytochromes, heme is bound covalently by way of two thioether bonds to the conserved CXXCH motif of the apoprotein in a posttranslational process referred to as cytochrome c maturation (*1*–*3*). Heme synthesis in the mitochondrial matrix (*4*) or bacterial cytoplasm (*2*), and the stereospecific, covalent heme attachment in the intermembrane space (*3*) or periplasm (*5*, *6*) are spatially separated processes requiring heme trafficking. Heme addition in mitochondria has been attributed to the enzyme cytochrome c heme lyase (*3*, *7*), although neither the mode of heme binding to that enzyme nor the mechanism of the ligation reaction has been elucidated. In *Escherichia coli*, eight *ccm* genes encode membrane proteins that are essential for cytochrome c maturation (*8*, *9*).

Escherichia coli genes *ccmABCDEFGH* were overexpressed from a plasmid to stimulate cytochrome c maturation. Analysis of the membrane fraction by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) revealed an 18-kD protein that retained peroxidase activity of c-type cytochromes with covalently bound heme. However, the size of this protein corresponded best to one of the products of the *ccm* genes, CcmE. A chromosomal in-frame deletion mutant, which was constructed by removing 92 *ccmE*-internal codons (Ile^3 to Ser^{94}) (*10*), was unable to produce mature c-type cytochromes (Fig. 1A). When *ccmE* was expressed in the Δ*ccmE* background from the arabinose-inducible promoter p_{ara} (*11*), membranes of the complementing strain contained both endogenous holocytochromes c and high levels of proposed heme-binding CcmE (Fig. 1A), as confirmed by immunoblot (Fig. 1B). The heme-protein association was SDS-resistant, as demonstrated by labeling *ccmE*-expressing *E. coli* cells with the heme precursor [^{14}C]-δ-aminolevulinic acid (δ-ALA) followed by SDS-PAGE of trichloroacetic acid (TCA)–precipitated cell extracts (*12*). Cells expressing the eight *ccm* genes plus the two naturally adjacent structural genes for the endogenous c-type cytochromes NapB and NapC on a multicopy plasmid (*6*) were transformed with a second plasmid containing an additional, arabinose-inducible *ccmE* gene and analyzed for heme-binding proteins (Fig. 1C). When *ccmE* expression from p_{ara} was repressed by the addition of glucose, endogenous *E. coli* c-type cytochromes such as NapB and NapC and the 18-kD CcmE protein were labeled to a similar extent. When *ccmE* was overexpressed by arabinose induction, however, most of the [^{14}C] label was incorporated into the 18-kD protein, confirming that CcmE contains a covalently bound tetrapyrrole. We conclude that the peroxidase activity (Fig. 1A) resulted from the presence of bound heme.

Next we characterized the spectroscopic features of the 18-kD hemoprotein. A hexa-

Mikrobiologisches Institut, Eidgenössische Technische Hochschule, Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland.

*To whom correspondence should be addressed. E-mail: lthoeny@micro.biol.ethz.ch