THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 266, No. 19, Issue of July 5, pp. 12242-12248, 1991
Printed in U.S.A.

Isolation and Deduced Amino Acid Sequence of the Gene Encoding gp115, a Yeast Glycophospholipid-anchored Protein Containing a Serine-rich Region*

(Received for publication, October 22, 1990)

Marina Vai, Evelina Gatti‡, Emanuela Lacanà§, Laura Popolo¶, and Lilia Alberghina

From Sezione di Biochimica Comparata, Dipartimento di Fisiologia e Biochimica Generali, Università di Milano, Via Celoria 26, 20133 Milano, Italy

gp115 is a *N*- and *O*-glycosylated protein of *Saccharomyces cerevisiae*. It is also modified by addition of glycosylphosphatidylinositol, which anchors the protein to the plasma membrane. The gene encoding gp115 (*GGP1*) has been cloned by a two-step procedure. By an immunoscreening of a yeast genomic DNA library in the expression vector λgt11, a 3′-terminal 0.9-kilobase portion of the gene has been isolated and then used as a molecular probe to screen a yeast genomic DNA library in YEp24. In this way, the whole *GGP1* gene has been cloned. Its identity with the gp115 gene has been confirmed by gene disruption, which has also indicated that the function of gp115 is not essential for cell viability. The features of the sequence are also entirely consistent with it corresponding to the gp115 gene. The nucleotide sequence of *GGP1* predicts a 60-kDa polypeptide, in agreement with the molecular mass of the gp115 precursor detected in *sec53* mutant cells at restrictive temperature. Two hydrophobic sequences, one NH_2- and the other COOH-terminal were found. The former has the features of the cleavable signal sequence, which allows the entry of proteins in the secretory pathway. The latter could be the signal sequence that has to be removed during the addition of glycosylphosphatidylinositol. The predicted amino acid sequence of gp115 shows 10 sequons for *N*-glycosylation and a high proportion of serine-threonine residues (22%) that could provide several sites for *O*-glycosylation. The unusual concentration of 27 serines in the COOH-terminal portion of the protein shares homology with a similar polyserine repeat of the serine repeat antigen (SERA protein) of *Plasmodium falciparum*. A two-dimensional analysis of the "*in vitro*" translational product of the *GGP1* mRNA has been carried out, allowing the identification of the "*in vivo*" gp115 precursor in a two-dimensional gel.

The addition of glycosylphosphatidylinositol (GPI)[1] is a recently discovered post-translational modification of proteins (1–3). The GPI is a plasma membrane anchoring structure and an increasing amount of evidence indicates that it constitutes the only membrane attachment for the GPI-linked proteins so far characterized. The GPI anchor is linked to the COOH terminus of the mature protein through an ethanolamine bridge (1, 2). Treatment with bacterial phosphatidylinositol-specific phospholipase C can release the GPI-anchored membrane protein in a soluble form. Studies are in progress to establish the physiological role played by the endogenous GPI-specific phospholipases in allowing this selective release (2, 4).

The GPI addition occurs in a number of proteins mainly of higher eukaryotic cells and protozoa. On the basis of their function, one can divide the GPI-linked proteins in several categories: cell surface enzymes, coat proteins or surface antigens, lymphoid antigens, and adhesion molecules (1, 2). Recent findings by our and other laboratories (5–7) indicate that GPI-linked proteins are also present in the yeast *Saccharomyces cerevisiae*, showing that this type of modification has been evolutionarily conserved among eukaryotic cells.

Metabolically labeled *myo*-[^{3}H]inositol proteins of *S. cerevisiae* have been resolved by high resolution two-dimensional gel electrophoresis (6). Among the four major species, we found gp115 (115 kDa, isoelectric point 4.8–5), which has been characterized as one of the more abundant plasma membrane glycoproteins (8, 9). The labeling of gp115 by palmitic acid also, its susceptibility to phosphatidylinositol-specific phospholipase C treatment, and its consequent behavior as a soluble protein in a detergent phase separation technique have indicated that this protein contains a GPI-like structure (6).

The synthesis of gp115 is highly regulated; it is active in growing cells and is shut off by conditions that cause yeast cells to arrest in stationary phase (10). For these reasons, the synthesis of this protein has been used as a marker of the exponential growth (11, 12). Besides, gp115 is a cell cycle modulated protein that reaches its maximal level between the G_1 and S phases (13, 14). Recently, we have further investigated the modulation of gp115 synthesis by using mutants of the cAMP metabolic pathway (15). We found that the synthesis of gp115 is triggered by cAMP in early START, the area of the G_1 phase in which central regulatory events of the cell cycle take place. The progression of the cell cycle toward DNA replication has been shown to require at this stage an increase of cAMP level and the activity of cAMP-dependent

* This work has been partially supported by a 1989 grant of the Ministero Pubblica Istruzione (40%) (to L. A.) and by P. F. Ingegneria Genetica, Consiglio Nazionale delle Ricerche (to L. P.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) X56399.

‡ Recipient of a fellowship from ENIRICERCHE.

§ Recipient of a fellowship from The Fondazione Anna Villa Rusconi.

¶ To whom correspondence should be addressed. Fax: 02-2361070.

[1] The abbreviations used are: GPI, glycosylphosphatidylinositol; ER, endoplasmic reticulum; kb, kilobase(s); SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis.

protein kinases (15). The addition of cAMP is shown to stimulate gp115 synthesis in cells arrested in early START even under conditions in which growth does not recover (10).

In this report, a final characterization of gp115 biosynthesis and the isolation and sequence of the gene (*GGP1*, glycophospholipid-anchored *g*lycosylated *p*rotein) encoding gp115 are presented. The deduced amino acid sequence is discussed.

EXPERIMENTAL PROCEDURES

Strains, Vectors, and Growth Condition—Escherichia coli strain Y1090 (*ΔlacU169 proA*$^+$ *Δlon araD139 strA SupF trp22::Tn10 pMC9*) was used for the screening of the yeast DNA library in λgt11. *E. coli* strain Y1089 (*ΔlacU169 proA*$^+$ *Δlon araD139 strA hflA chr::Tn10 pMC9*) was lysogenized with the recombinant phages and used for protein analysis and for phage DNA preparations. Both these strains, originally isolated by Young and Davis (16), were obtained from G. Lucchini (Dipartimento di Genetica e Biologia dei Microorganismi, Università di Milano, Italy). Growth media and phage infection conditions were as described by Sambrook *et al.* (17). The yeast genomic DNA library in λgt11 (16) was kindly provided by R. W. Davis (Stanford University, Stanford, CA). *E. coli* JM101 (*Δ(lac pro) thi strA supE endA sbcB hsdR F′traD36 proAB lacI9 lacZ ΔM15*) was the host strain for the transformation with pGEM plasmids obtained from Promega Biotec. *E. coli* HB101 (*F*$^-$*hsd20 (rB*$^-$*,mB*$^-$*) supE44 ara-14 galK-2 lacY-1 proA2 rpsL20 (str*^{r}*) XY1-5 mtl-1 recA13*) was the host strain of the yeast genomic DNA library in the YEp24 plasmid described in (18). This DNA library was kindly provided by G. Lucchini. Plasmid pVGLP was obtained by cloning the 5-kb *Nco*I yeast DNA fragment (Fig. 2*A*) in the *Nco*I restriction site of the pGEM-5Zf(+) polylinker. Plasmid YEpBS6 was constructed by cloning the 4.9-kb *Bgl*II-*Sph*I fragment of pVGLP into the *Bam*HI- and *Sph*I-digested YEp24. Plasmid pXH was obtained by cloning the 1.9-kb *Xho*I-*Hin*dIII fragment (Fig. 2*A*) into the *Xho*I- and *Hin*dIII-digested pGEM-7Zf(+).

S. cerevisiae secretory mutant strain SFNY28-12B (*MATα, sec53, his3-11*) was obtained by P. Novick (Yale University), the haploid GRF18 strain used for gene disruption (*MATα, his3-11,15, leu2-3,112*) was kindly provided by A. Hinnen (CIBA-Geigy, Basel, Switzerland), and the haploid W303-1B strain (*MATα, ade2-1, can1-100, ura3-1, leu2-3,112, trip1-1, his3-11,15*) isolated by R. Rothstein, was supplied by P. P. Slonmiski (Centre National de la Recherche Scientifique, Gift-sur-Yvette, France). The wild type strain S288C (*MATα, SUC2, mal, mel, gal2, CUP1*) was obtained from the Yeast Genetic Stock Center (Berkeley, CA), and the strain 6075/5C (*MAT***a**, *leu2-1*) was kindly provided by M. L. Agostoni (Dipartimento di Genetica e Biologia dei Microorganismi, Università di Milano, Italy).

Cells were grown in batches in YEPD medium containing 1% yeast extract, 2% Bacto-peptone, and 2% glucose, or in DIFCO yeast nitrogen base (YNB) medium (6.7 g/liter), 2% glucose and supplemented with the required amino acids. SFNY28-12B cells were grown at 25 °C (permissive temperature) while all the other strains were grown at 30 °C.

Screening Procedures—A procedure slightly modified from Young and Davis (16) was used for immunoscreening of recombinant phages (19). Before use, anti-*E. coli* antibodies were removed from the antiserum anti-gp115 as described in Ref. 19.

Immuno-positive clones were characterized by restriction endonuclease mapping, by the size of the hybrid proteins produced in lysogenes (6) and by DNA hybridization analysis.

The screening of the DNA genomic library constructed by size fractionation of partially digested DNA in YEp24 (18) was performed according to standard techniques (17). About 16,000 clones, representing more than 99% of the yeast genome (20) were screened using a ^{32}P-labeled RNA probe prepared as described below.

DNA and RNA Analysis—Standard protocols, including cloning procedures, plasmid DNA preparation, bacterial transformation, agarose or acrylamide gel electrophoresis, and Southern and Northern analyses have been described by Sambrook *et al.* (17). λgt11 DNA and recombinant phage DNA were purified using the Promega LambdaSorb™ phage adsorbent. Transformation of *S. cerevisiae* cells was as described by Ito *et al.* (21). Total yeast DNA was prepared by the method of Nasmyth and Reed (22) and yeast RNA according to Elion and Warner (23). Poly(A)$^+$ RNA was purified by affinity chromatography on Hybond-mAP paper (Amersham Corp.).

Radiolabeled single-strand RNA probes were generated by "*in vitro*" transcription (24) of DNA fragments inserted into the multiple cloning site of pGEM-blue plasmid using the Riboprobe™ system from Promega Biotec and [α-^{32}P]UTP (800 Ci/mmol, Amersham Corp.).

Gene Disruption—To determine if the cloned fragment contained the *GGP1* gene, the procedure of one-step gene disruption described by Rothstein (25) has been used. The plasmid pVGLP was digested with *Xho*I and *Sal*I, and the selectable *LEU2* yeast gene, derived from YEp13 vector (26), was cloned into the cleaved DNA fragment. *E. coli* JM101 cells were transformed, and plasmid DNA analyzed by restriction mapping to confirm the correct insertion of the fragment. The recombinant plasmid was digested with *Nco*I in order to excise the yeast DNA fragment from the vector, and the whole mixture was used to transform GRF18 Leu$^-$ cells. Disruptions were tested by Southern analysis to confirm the gene replacement.

DNA Sequencing and Nucleotide Sequence Analysis—Two strategies were principally used. (i) Different restriction fragments of the 5-kb *Nco*I-fragment were subcloned. (ii) Deletions were obtained using *Exo*III according to Henikoff (27) or using *Bal*31 nuclease (17). Plasmid DNA was purified on CsCl gradients. In some instances, 18-base oligonucleotides were synthesized by a DNA synthesizer (Cyclone, Biosearch, Inc.) and used as sequencing primers. Sequencing on both strands was done by the dideoxy chain termination method (28) using alkaline denatured DNA templates. The sequencing reactions were carried out with the Pharmacia ^{T7}DNA polymerase sequencing kit using the [^{35}S]dATP as the labeled nucleotide. Promega TaqTrack™ sequencing system was used to resolve the ambiguities eventually found in the sequence.

The nucleotide sequence of *GGP1* and the deduced amino acid sequence of gp115 were analyzed using the programs of the PCGENE package from IntelliGenetics.

Extract Preparation, Labeling Conditions, Electrophoretic Procedures, and Immunoblotting—Protein extracts from *E. coli* lysogenes were prepared according to the procedure described by Young and Davis (16). Total unlabeled protein extracts of *S. cerevisiae* cells were prepared as previously described (29). [^{35}S]Methionine (20 μCi/ml) was added to *sec53* exponentially growing cells or 15 min after the shift to 37 °C, as previously reported (8). After 1.5 h, labeling extracts were prepared for the analysis by two-dimensional gel electrophoresis (14, 30).

Proteins were resolved by SDS-PAGE (31) or by two-dimensional gel electrophoresis (30) on 8% polyacrylamide slab gels. Labeled gels, following electrophoresis, were treated with Enlightning (Du Pont-New England Nuclear). Electrophoretic transfer of proteins and immunodecoration of blots were carried out as previously described (32, 33).

In Vitro Transcription and Translation—About 1–2 μg of *Hin*dIII-linearized plasmid pXH was transcribed *in vitro* according to a procedure previously reported (34). The transcription reaction (40 μl) was carried out with T7 RNA polymerase in the presence of m^{7}G(5′)ppp(5′)G for capping at 37 °C for 1.5 h. One-twentieth of the transcription product was translated in a rabbit reticulocyte lysate (Promega Biotec) according to the conditions suggested by the manufacturer with 1 μCi/ul of [^{35}S]methionine for 1 h at 30 °C. Aliquots of the reaction mixture were analyzed by two-dimensional gel electrophoresis.

RESULTS

Biosynthetic Pathway of gp115—In previous papers from our laboratory, a partial characterization of the gp115 biosynthesis has been described (6, 9). By using tunicamycin, an inhibitor of *N*-glycosylation but not of GPI addition, an 88-kDa polypeptide (p88) is detected (6, 35). *N*-Core oligosaccharide chains linked to the protein contribute to the formation of a 100-kDa intermediate precursor (p100). Transport from endoplasmic reticulum (ER) and final maturation of p100 into gp115 takes places in the Golgi and can be blocked by the *sec18* temperature-sensitive mutation (6). Here, we used the *sec53* mutant to study the early steps of gp115 biosynthesis and, in particular, whether *O*-glycosylation occurs in gp115. The SEC53 gene is known to encode the cytoplasmic enzyme phosphomannomutase that is required for the production of GDP-mannose (36). It has been shown only very recently that this sugar donor is required not only for the *N*- and *O*-glycosylation, but also for the formation of

the glycans moiety of the GPI structure (7). In order to roughly estimate the contribution of the *O*-linked mannoses and of GPI to the mass of the p88 *N*-unglycosylated form, SDS-PAGE-fractionated protein extracts from *sec53* cells grown at 25 °C or incubated for 2 h at 37 °C were analyzed by an immunoblot using anti-gp115 antibodies. A 65-kDa immunoreactive band (p65) was accumulated at 37 °C (Fig. 1*A*, *lane b*). The gp115 accumulated before the temperature shift-up was also detected since this protein is rather stable (about 10% of degradation/h) (data not shown). Since GPI contributes only for about 2–4 kDa of the apparent molecular mass of a polypeptide (1,6), it appears that gp115 is also modified by *O*-glycosylation. A schematic representation of gp115 biosynthesis is shown in Fig. 1*B*.

Isolation of the Gene (GGP1) Encoding gp115—We have screened a genomic yeast DNA library in the expression vector λgt11 (16) (2×10^5 recombinant phages) using antibodies anti-gp115 previously obtained (9). Phages from 135 plaques giving positive signals at the initial screening were replated and rescreened several times. The last immunoscreening of 12 recombinant clones was performed with antibodies affinity-purified from p88 (6, 9). One recombinant phage (λ7) containing a 2.7-kb *Eco*RI yeast DNA fragment (Fig. 2*A*) inserted in the LacZ gene and producing a 137-kDa hybrid protein was further analyzed. A 0.9-kb *Eco*RI-*Hin*dIII fragment contiguous to the 3′ end of the LacZ gene was sequenced and it was shown to contain a coding region in frame with LacZ and a stop translation codon close to the *Hin*dIII restriction site, thus making likely that it could encode the 23-kDa heterologous polypeptide moiety of the 137-kDa hybrid product, 114 kDa being the molecular mass of β-galactosidase in the fusion protein (19).

The 0.9-kb *Eco*RI-*Hin*dIII fragment found in λ7 (Fig. 2*A*) was used as a probe in Southern analysis of total yeast DNA (Fig. 2*B*), allowing the construction of a preliminary restriction map of the yeast cloned genomic locus (Fig. 2*A*). To isolate the entire gp115 gene, we used this fragment as a molecular probe to screen a yeast genomic DNA library in YEp24 plasmid (18). All the 17 recombinant clones selected after several hybridization steps contained genomic inserts of about 10 kb with an internal 5-kb *Nco*I fragment common to all the classes of plasmids. This internal fragment (Fig. 2*A*), cloned in the high copy number vector YEp24 (YEpBS6

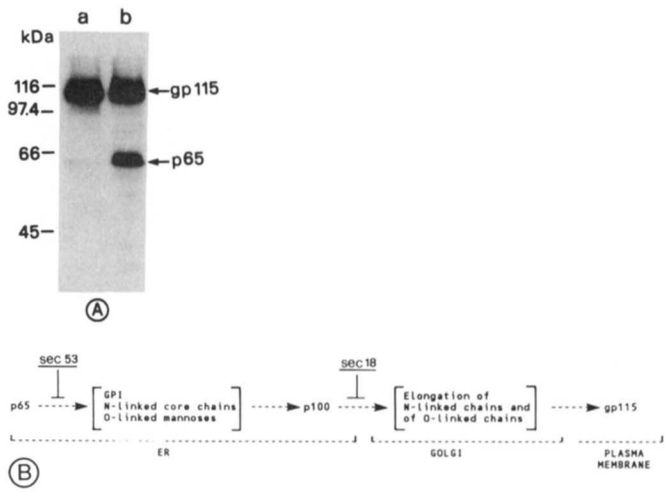

FIG. 1. **Biosynthetic pathway of gp115.** *A*, total extracts from exponentially growing *sec53* cells at 25 °C (*lane a*) and after 2 h incubation at 37 °C (*lane b*) were fractionated by SDS-PAGE. Blots were probed with anti-gp115 antiserum and ^{125}I-labeled protein A. *B*, schematic representation of gp115 biosynthesis.

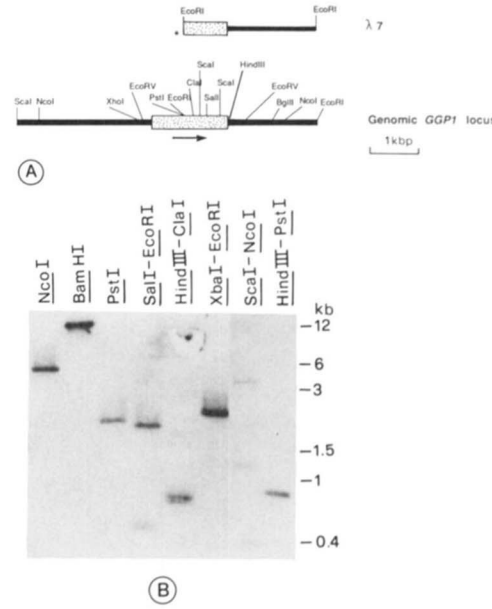

FIG. 2. **Isolation of *GGP1* gene.** *A*, structure of the yeast DNA insert in λ7 recombinant phage (*upper panel*), with an *asterisk* indicating the point of fusion with the 3′-end of the LacZ gene (not shown). *Lower panel*, restriction map of yeast genomic DNA. The *arrow* indicates the direction of transcription. The *shaded boxes* in both *panels* indicate the GGP1-coding region. *B*, Southern analysis of restricted genomic DNA probed with the 0.9-kb *Eco*RI-*Hin*dIII fragment from λ7.

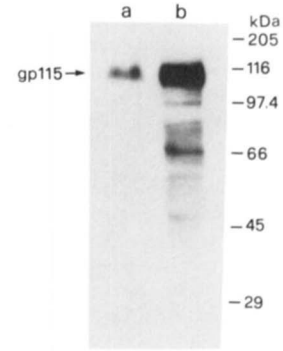

FIG. 3. **Overexpression of gp115.** Total extracts from exponentially growing W303-1B cells (*lane a*) and from cells transformed with the multicopy plasmid YEpBS6 (*lane b*), whose construction is described under "Experimental Procedures," were fractionated by SDS-PAGE, blotted on nitrocellulose, and immunodecorated as in Fig. 1*A*. The same amount of proteins was loaded on each lane.

plasmid, see "Experimental Procedures"), was used to transform the haploid strain W303-1B. The transformed Ura$^+$ cells were shown to contain the YEpBS6 plasmid by Southern analysis (data not shown) and to overproduce the gp115 about 5-fold over the control cells (Fig. 3, *lanes a* and *b*), thus indicating that the complete gene is contained in the 5-kb *Nco*I fragment. In Fig. 3, *lane b*, besides gp115, other immunoreactive bands of lower molecular weight were present, probably due to gp115 degradation or to partially processed forms of gp115.

mRNA Analysis—Northern analysis of yeast poly(A)$^+$ RNA was carried out. Two RNA probes complementary to the 0.9-kb *Eco*RI-*Hin*dIII fragment (see Fig. 2*A*) and transcribed in two different orientations allowed us to detect a 1.9-kb mRNA (Fig. 4), which is transcribed in the *Eco*RI-*Hin*dIII direction according to the orientation of the yeast DNA insert found in λ7 (Fig. 2*A*). The size of mRNA is compatible with the

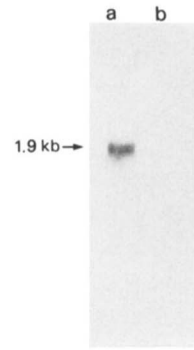

FIG. 4. **Northern blot analysis of yeast poly(A)$^+$ RNA.** Identical samples of about 20 μg of poly(A)$^+$ RNA from strain S288C prepared as described under "Experimental Procedures" were hybridized with complementary single-strand ^{32}P-labeled probes. These were prepared by transcribing with SP6 (*lane a*) or with T7 (*lane b*) RNA polymerases the 0.9-kb *Eco*RI-*Hin*dIII fragment cloned into the *Eco*RI- and *Hin*dIII-digested pGEM-blue vector.

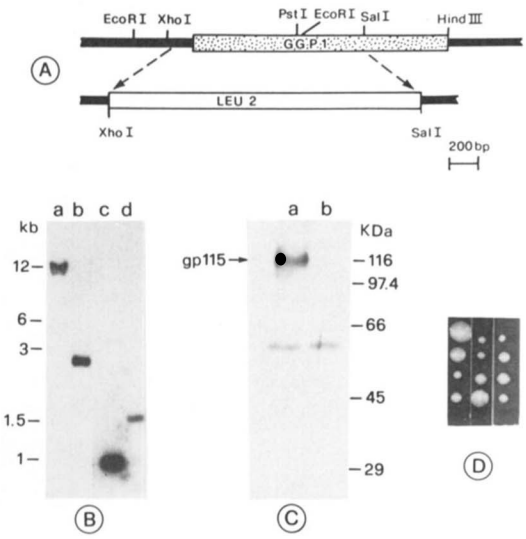

FIG. 5. **Disruption of the *GGP1* gene.** *A*, schematically shown is the replacement of the *Xho*I-*Sal*I fragment from the *GGP1* gene with a 2.2-kb *Xho*I-*Sal*I fragment harboring the *S. cerevisiae* LEU2 gene. *B*, Southern analysis of restricted DNA of GRF18 strain (*lanes b* and *c*) and of a Leu$^+$ transformant (*lanes a* and *d*) probed with the 0.9-kb *Eco*RI-*Hin*dIII fragment. Yeast DNA was digested with *Pst*I (*lanes a* and *b*) or with *Eco*RI and *Hin*dIII (*lanes c* and *d*). *C*, total protein extracts of exponentially growing GRF18 cells (*lane a*) and of the Leu$^+$ transformant (*lane b*) were fractionated by SDS-PAGE and analyzed by immunoblot as described in Fig. 1*A*. *D*, meiotic progeny from dissected tetrads of diploids heterozygous for the *ggp1*::LEU2 disruption is shown.

apparent molecular mass of the precursor of gp115 (65 kDa).

Confirmation of Gene Identity—To unambiguously ascertain whether the *GGP1* gene is contained in the cloned region, a null mutation of the *GGP1* locus was generated in the haploid strain GRF18 by the one-step gene disruption procedure using *LEU2* as a selective marker (Fig. 5*A*). The *ggp1*::*LEU2* disruption construct is described under "Experimental Procedures." Southern analysis of one of the Leu2$^+$ transformants, showing correct chromosomal integration of the disrupting fragment is shown in Fig. 5*B*. The hybridization with the 0.9-kb *Eco*RI-*Hin*dIII shows that the expected 2.4-kb *Pst*I and 0.9-kb *Eco*RI-*Hin*dIII fragments are detected only in the wild type cells (*lanes b* and *c*), whereas bands of higher molecular weight, due to the integration of LEU2, are present in the disrupted cells (*lanes a* and *d*).

Total protein extracts of these transformant cells were analyzed by immunoblot. As shown in Fig. 5*C*, *lane b*, neither gp115 nor a truncated form of it are produced by the disrupted strain. The lower molecular weight band present both in the control and the disrupted strains is due to a cross-reactive protein that is detected when anti-gp115 antiserum instead of affinity p88-purified antibodies is used (9).

The ability of haploid mutant cells to grow indicates that this gene is not essential. Nevertheless, *ggp1* disruptants show severe morphological aberrations. Cells are larger than wild type ones, round-shaped, often contain more than one bud even when stationary phase has been reached, and thus appear to be defective in cell separation (data not shown). The duplication time is from 20% up to 30% lower than that of the isogenic strain, and cellular viability during stationary phase is greatly reduced.

The haploid null mutant cells (*ggp1*::*LEU2*) were crossed with the wild type strain 6075/5C. Diploid cells were sporulated, and segregants were examined by tetrad analysis. All tetrads gave rise to four viable spores with a 2:2 segregation of Leu$^+$:Leu$^-$ colonies, the smaller of which were of Leu2$^+$ phenotype (Fig. 5*D*). The morphological aberrations determined by *GGP1* gene disruption cosegregated exactly with the Leu$^+$ phenotype, whereas the Leu$^-$ one was always associated to a wild type morphology. This provides further confirmation that integration has occurred at the *GGP1* locus.

Sequencing and Structural Analysis of the GGP1 Gene—The sequence of the *GGP1* gene and of its flanking regions has been obtained as described under "Experimental Procedures." The sequence contains a single extended open reading frame that initiates with the ATG at nucleotide 1 and spans to the TAA at nucleotide 1677 (Fig. 6). In this region, the sequence is 58% A+T, whereas the nucleotide sequences between −300 to −1 and +1680 to +1996 are 67 and 72%, respectively. Such AT-rich noncoding sequences have been found adjacent to most of the yeast genes already sequenced (37). In the region 5′ upstream the reading frame several sequences similar to the canonical TATA box are present, as is typical of yeast genes. The first methionine codon in the reading frame is surrounded by a sequence that is a six-out-of-nine match with the yeast-optimal AUG start codon context (38). Moreover, the region 3′ downstream of the gene contains elements similar to the consensus sequence proposed by Zaret and Sherman (39) to be involved in transcription termination in yeast. The open reading frame would encode a protein of 559 amino acids with a molecular weight of 59,581 Da and a predicted isoelectric point (pI) of 4.14. Several features associated with the *GGP1* open reading frame are consistent with the expected properties of the product of the gp115 gene (Figs. 6 and 7). At the amino terminus of the open reading frame, a hydrophobic sequence spanning from amino acid 2 to 22 is present (LFKSLSKLATAAFFAGVATA) (Figs. 6 and 7*A*). According to the method of MohanaRao and Argos (40), it can be predicted that this sequence should give origin to a transmembrane α-helix. This prediction is also in agreement with results obtained with the method of Eisenberg *et al.* (41). Other additional features, such as the presence of a hydrophobic core and of a potential cleavage site after amino acid residue 22, obeying the rule "minus 3 minus 1" (42) with both residues being alanine, make this sequence similar to the signal sequences of many secretory proteins of prokaryotic and eukaryotic cells. The presence of a putative signal sequence is in agreement with the fact that gp115 is transported to the plasma membrane through the secretory pathway (6, 9). The size of the putative product of the *GGP1* gene is fairly consistent with the size of the gp115 protein, of which about 44% is constituted by *N*- and *O*-linked carbohydrates and

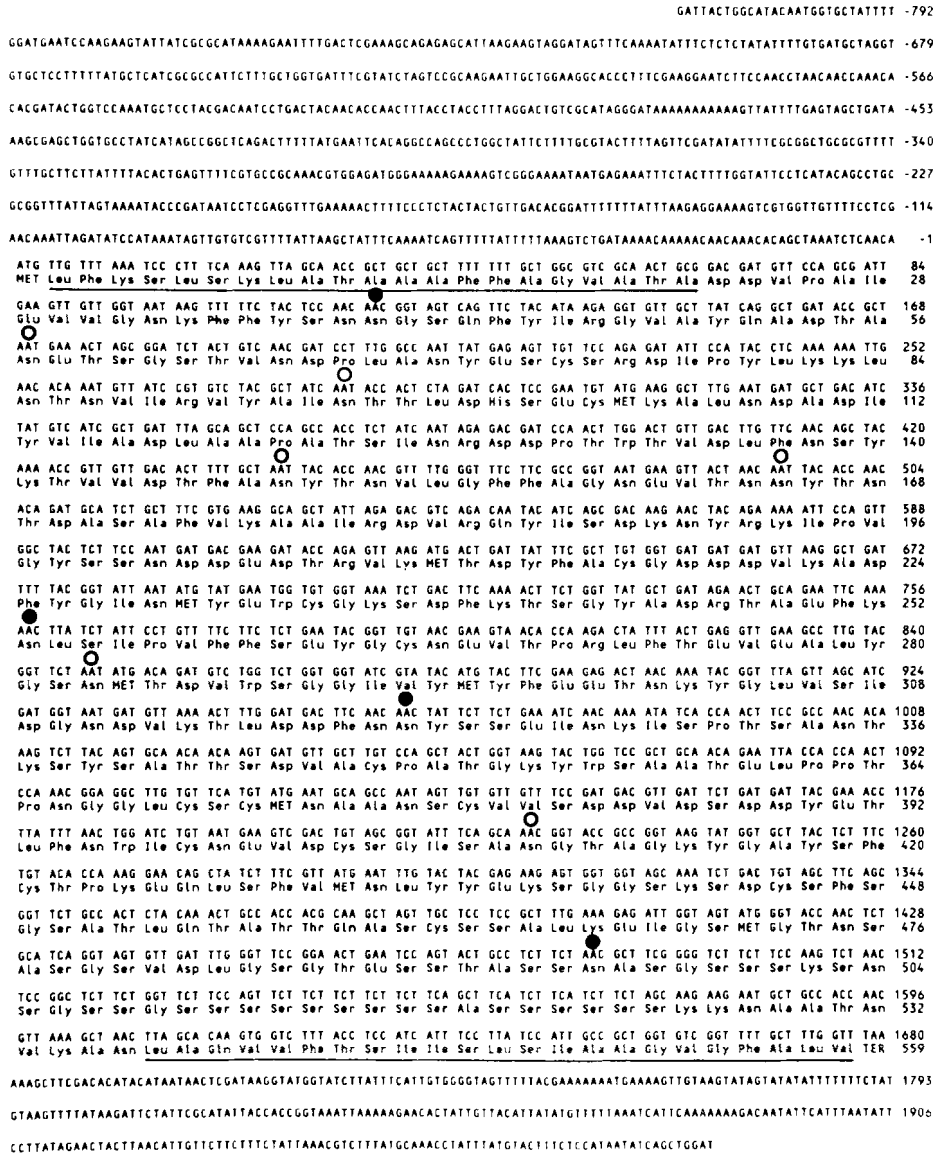

FIG. 6. **Sequence of the *GGP1* gene.** The complete sequence of the *GGP1* gene, as well as 819 and 316 nucleotides of its 5′- and 3′-flanking regions are presented. The predicted gp115 open reading frame starting at nucleotide 1 is shown under the corresponding nucleotide sequence. The NH_2- and COOH-terminal hydrophobic sequences are *underlined*. The potential *N*-glycosylation sites, Asp-*X*-Ser (●) and Asp-*X*-Thr (○), are shown.

GPI. The putative gp115 protein has 10 potential *N*-glycosylation sites (Asn-*X*-Ser or Thr) and a high proportion of serines and threonines (22%) (Fig. 6). The distribution of serines and threonines is shown in Fig. 7*B*. The profile shows the presence of a remarkable region in the COOH-terminal portion of the gene in which the serine residues are particularly concentrated. The serine and threonine residues could provide possible sites for *O*-glycosylation.

Overall, the high content of hydroxyamino acids (Ser, Thr, and Tyr = 28%) can account for the slight discrepancy between the predicted molecular weight of gp115 (59.85 kDa) and the apparent molecular mass of 65 kDa of its precursor, as determined by SDS-PAGE (Fig. 1*A*, *lane b*), as described for other proteins (43, 44).

The last 22 amino acids of the derived protein sequence (from position 537 to 558) also contain a hydrophobic sequence (LAQVVFTSIISLSIAAGVGFAL) (Figs. 6 and 7*A*) for which the formation of a transmembrane α-helix is predicted by two different methods of analysis (40, 41). The presence of a COOH-terminal hydrophobic sequence is common to all the gene products so far characterized that undergo GPI addition after translation (1, 35). A similar sequence has been found in the COOH-terminal of the yeast cell surface α-agglutinin and *KRE1* gene product, therefore suggesting that these are also GPI-linked proteins (44, 45). In the case of gp115, the presence of the putative recognition sequence required for GPI addition strengthens the previous findings indicating that this protein contains GPI (6).

A homology search of the protein data banks SWISS-PROT and NBRF showed no significant homology of this protein to any other. The protein with the highest match score is the serine repeat antigen protein precursor, encoded by the *SERA* gene, of *Plasmodium falciparum* (46). The match between 26 out of the 37 amino acids of the serine repeat of the SERA protein and a serine-rich sequence of gp115 (see Fig. 7*C*) accounts almost entirely for the similarity between the two proteins. A rather scattered similarity has also been found with some cell surface glycoproteins of *S. cerevisiae* (for example the a- and α-agglutinins (45, 47) and the *KRE1* gene product (44)). Also in these cases, the similarities are restricted to the matches between the serines. The 27 serines present in the serine-rich sequence of the gp115 are preferentially coded by the TC$\binom{T}{C}$ codons (48).

Two-dimensional Electrophoretic Analysis of in Vitro Translation of GGP1 mRNA—The *Hin*dIII-linearized plasmid pXH

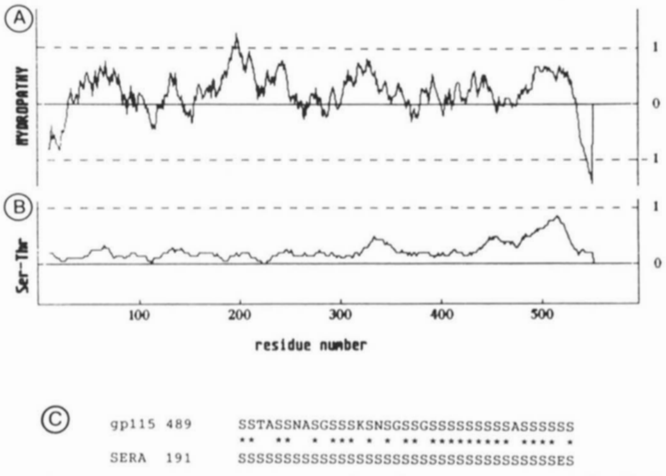

FIG. 7. **Structural features of gp115.** *A*, Kyte and Doolittle analysis (52) for hydrophilic and hydrophobic regions with a window of 21 amino acids. The scale of Kyte and Doolittle was normalized according to Hopp (53). *B*, frequency of Ser and Thr residues using a value of 1 for Ser and Thr and a value of 0 for all other amino acids with a window of 21 amino acids. *C*, alignment between sequences of serine-rich regions of gp115 and of the SERA protein of *P. falciparum*. The *asterisk* indicates the position of the conserved amino acids.

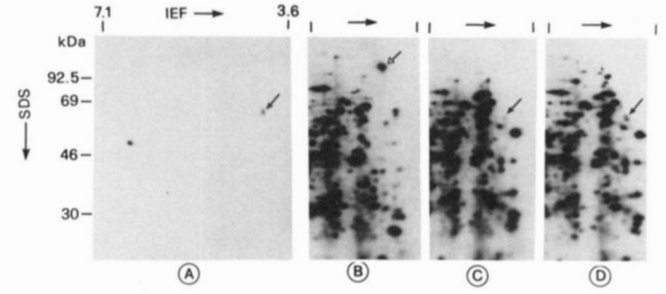

FIG. 8. **Two-dimensional analysis of the *in vitro* and *in vivo* labeled p65.** *A*, a rabbit reticulocyte lysate supplemented with [^{35}S]methionine was incubated with the *in vitro* transcript from *Hin*dIII-linearized pXH, and the translational products were analyzed by two-dimensional gel electrophoresis. The *closed arrow* indicates the translational product of *GGP1* mRNA. *B*, exponentially growing *sec53* cells labeled at 25 °C with [^{35}S]methionine. The *open arrow* indicates the gp115. *C*, *sec53* cells shifted to 37 °C and labeled 15 min later with [^{35}S]methionine. The *in vivo* p65 is indicated by the *closed arrow*. *D*, a total labeled extract obtained as in *C* was mixed with the *in vitro* translational product. The *closed arrow* indicates the comigration of the two spots. Acidic sections of two-dimensional gels are shown in *B–D*. The pH range is from 5.5 (*left*) to 3.6 (*right*). The same amount of disintegrations/min was loaded on each of the gels of *B–D*. *IEF*, isoelectric focusing.

was transcribed *in vitro* with the T7 RNA polymerase, as described under "Experimental Procedures." The synthetic mRNA obtained was then translated *in vitro* in a rabbit reticulocyte lysate, and the translational products were analyzed by two-dimensional gel electrophoresis. As shown in Fig. 8*A*, the *GGP1* mRNA is efficiently translated *in vitro* in a polypeptide of about 62–65 kDa with a pI of 4.1, perfectly in agreement with the molecular mass of the gp115 precursor (see Fig. 1*A*, *lane b*) and with the theoretical pI 4.14 calculated from the *GGP1* open reading frame. The presence of another basic protein of about 50 kDa in Fig. 8*A* can be ascribed to the *in vitro* translation of the globin mRNA that is present in the reticulocyte lysate used.

The availability of the *in vitro* translational product of *GGP1* mRNA allowed us also to identify the "*in vivo*" gp115 precursor (see Fig. 1*A*, *lane b*) in a two-dimensional gel. For this reason, *sec53* cells were labeled both at 25 and 37 °C with [^{35}S]methionine, and at the end of the labeling period (1.5 h), proteins were subjected to two-dimensional gel electrophoresis (Fig. 8, *B* and *C*). The labeled extract obtained at 37 °C was also comigrated with the *in vitro* translational product, and the two-dimensional fluorogram is shown in Fig. 8*D*. At 25 °C, gp115 (115 kDa and pI 4.8–5) is clearly recognizable, whereas at 37 °C, it is no longer detected as labeled protein since the [^{35}S]methionine added 15 min after the shift to the restrictive temperature is incorporated only in the proteins synthesized during the *sec53* block. Moreover, by comparison of the patterns shown in Fig. 8, *C* and *D*, one can identify the *in vivo* gp115 precursor (p65) that on two-dimensional gel comigrates with the *in vitro* synthesized one. This precursor is specifically accumulated in *sec53* at 37 °C, but its level appears lower than expected on the basis of SDS-PAGE analysis. This is probably due to a very poor solubilization of this protein in the urea lysis buffer used for two-dimensional gel electrophoresis, as reported for other precursors blocked at the ER level (49).

DISCUSSION

In this paper, we describe the isolation and the molecular characterization of the gp115 structural gene, *GGP1*. The results of Southern analysis and gene disruption suggest that this gene is present as a single-copy gene in the yeast genome and it is not essential for cell viability, although *ggp1* disruptants show morphological aberrations. The predicted amino acid sequence analysis of *GGP1* allows the deduction and confirmation of relevant structural features of gp115. Consistent with the fact that gp115 is one of the major yeast glycoproteins, the *GGP1* open reading frame contains 10 consensus sequences (sequons) for *N*-glycosylation. Previous studies have indicated that approximately six to seven core *N*-linked chains are attached to gp115 (6, 9). This implies that not all the sequons are actually modified, probably due to steric hindrance to the modifying enzymes. In this paper, we show that gp115 also undergoes *O*-glycosylation. Haselbeck and Tanner (50) have shown that, in yeast, *O*-glycosylation begins with the addition of 1 mannose residue to serine or threonine in the ER and it is completed in the Golgi by the addition of other 3–4 mannose residues. In agreement with the finding that gp115 is highly *O*-glycosylated, a high proportion of serines and threonines has been found. Serine is about 14.6% and threonine is about 8% of the total amino acids present in gp115, as compared with an average value of about 7 and 5.7%, respectively, in all the yeast proteins present in the Protein Identification Resourse database (51). In particular, a total of 27 serine residues is concentrated in a 37-amino acid-long stretch inside of which two clusters of 9 and 6 consecutive serines are present (Fig. 7*C*). This is a rather unusual structural feature for yeast cell wall mannoproteins. In fact, a high proportion of serine and threonine is known to be common to this kind of protein, forming in some cases a serine-threonine-rich region (45). However, in none of the so far characterized mannoproteins has such a high number of consecutive serines been found. We propose that, in gp115, most of the *O*-linked carbohydrates are in this portion of the protein and that such a modified serine-rich domain assumes an extended and stable conformation. Since this structure is localized in the COOH-terminal portion of the protein, one can tentatively suppose that it could function as a spacer between the GPI-anchoring site in the plasma membrane and the exterior of the cell wall. The great concentration of ordered negative charges brought by the clustered carbohydrates in such an abundantly present protein could play a role in keeping positive ions in close proximity of the

plasma membrane. This hypothesis, as well as the possible structural role of gp115 in the microarchitecture of the cell wall, is currently under investigation in our laboratory. For the homologous polyserine structure of the SERA protein of *P. falciparum*, no biochemical function has yet been proposed (46). The SERA protein is a blood stage antigen able to induce parasite-inhibitory antibodies (46). It has been recently proposed that it could function at the red blood cell membrane stage as a component influencing the invasion process. In agreement with the observation that this antigen is exported, a potential leader sequence has been found in the amino acid sequence predicted from the cloned SERA gene (46). *N*-Glycosylation sites and a high proportion of serine and threonine residues are also present. Also, in this case, the possibility that *O*-glycosylation could occur at any of many possible sites has been proposed (46). The advantages of yeast as an experimental system for the analysis of cloned gene products offer the possibility of exploiting the structural features of gp115 to gain insight into the function of serine-rich regions.

The finding that a hydrophobic sequence is present in the COOH terminus of the predicted gene product is consistent with the addition of GPI to gp115. In fact, it is known that GPI-linked proteins are synthesized as precursors with the hydrophobic carboxyl termini (1). The linkage of the GPI occurs in the ER and involves the removal of the hydrophobic sequence and the covalent attachment of the carboxyl-terminal amino acid of the mature protein to the anchor. The recent findings of similar terminal sequences in *AGα1* and *KRE1* genes, respectively encoding for the cell surface α-agglutinin (45) and for an enzyme involved in the production of cell wall (1→6)-β-glucan (44), suggest that in yeast also the GPI-linked proteins exhibit different physiological functions. The elucidation of the structure-to-function relationship in gp115 is the object of our present studies.

Acknowledgments—We wish to express our thanks to R. Grandori, S. Livian, and F. Rodriguez for technical assistance in immunoscreening and sequencing and to A. Grippo for the elaboration and preparation of the figures. We are grateful to G. Grandi and A. Tognoni of ENIRICERCHE, San Donato Milanese, for the facility and assistance in the computer analysis and to P. Pedroni for the preparation of oligonucleotides.

REFERENCES

1. Ferguson, M. A. J., and Williams, A. F. (1988) *Annu. Rev. Biochem.* **57,** 285–320
2. Low, M. G. (1989) *Biochim. Biophys. Acta* **988,** 427–454
3. Low, M. G., and Saltiel, A. R. (1988) *Science* **239,** 268–275
4. Braun-Breton, C., Rosenberry, T. L., and Pereira da Silva, R. (1988) *Nature* **332,** 457–459
5. Conzelmann, A., Riezman, H., Desponds, C. D., and Bron, C. (1988) *EMBO J.* **7,** 2233–2240
6. Vai, M., Popolo, L., Grandori, R., Lacanà, E., and Alberghina, L. (1990) *Biochim. Biophys. Acta* **1038,** 277–285
7. Conzelmann, A., Fankhauser, C., and Desponds, C. (1990) *EMBO J.* **9,** 653–661
8. Popolo, L., Vai, M., and Alberghina, L. (1986) *J. Biol. Chem.* **261,** 3479–3482
9. Popolo, L., Grandori, R., Vai, M., Lacanà, E., and Alberghina, L. (1988) *Eur. J. Cell Biol.* **47,** 173–180
10. Grandori, R., Popolo, L., Vai, M., and Alberghina, L. (1990) *J. Biol. Chem.* **265,** 14315–14320
11. Verma, R., Iida, H., and Pardee, A. B. (1988) *J. Biol. Chem.* **263,** 8569–8575
12. Verma, R., Iida, H., and Pardee, A. B. (1988) *J. Biol. Chem.* **263,** 8576–8582
13. Lorincz, A. T., Miller, M. J., Xuong, N. H., and Geidushek, E. P. (1982) *Mol. Cell. Biol.* **2,** 1532–1549
14. Popolo, L., and Alberghina, L. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 120–124
15. Matsumoto, K., Uno, I., Kato, K., and Ishikawa, T. (1985) *Yeast* **1,** 25–36
16. Young, R. A., and Davis, R. W. (1983) *Science* **222,** 778–782
17. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
18. Carlson, M., and Botstein, D. (1982) *Cell* **28,** 145–154
19. Lucchini, G., Brandazza, A., Badaracco, G., Bianchi, M., and Plevani, P. (1985) *Curr. Genet.* **10,** 245–252
20. Clarke, L., and Carbon, J. (1976) *Cell* **9,** 91–94
21. Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983) *J. Bacteriol.* **153,** 163–168
22. Nasmyth, K. A., and Reed, S. I. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 2119–2123
23. Elion, E. A., and Warner, J. R. (1984) *Cell* **39,** 663–673
24. Melton, D. A., Krieg, P. A., Relagliati, M. R., Maniatis, T., Zinn, K., and Green, M. R. (1984) *Nucleic Acids Res.* **12,** 7035–7056
25. Rothstein, R. J. (1983) *Methods Enzymol.* **101,** 202–211
26. Broach, J. R., Strathern, J. N., and Hicks, J. B. (1979) *Gene (Amst.)* **8,** 121–133
27. Henikoff, S. (1984) *Gene (Amst.)* **28,** 351–359
28. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5413–5417
29. Vai, M., Popolo, L., and Alberghina, L. (1986) *FEBS Lett.* **206,** 135–141
30. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250,** 4007–4021
31. Laemmli, U. K. (1970) *Nature* **227,** 680–685
32. Towbin, H., Staehelin, T., and Gordon, J. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76,** 4350–4354
33. Zippel, R., Sturani, E., Toschi, L., Naldini, L., Alberghina, L., and Comoglio, P. M. (1986) *Biochim. Biophys. Acta* **881,** 54–61
34. Krieg, P. A., and Melton, D. A. (1984) *Nucleic Acids Res.* **12,** 7057–7070
35. Cross, G. A. M. (1987) *Cell* **48,** 179–181
36. Kepes, F., and Schekman, R. (1988) *J. Biol. Chem.* **263,** 9155–9161
37. Taussig, R., and Carlson, M. (1983) *Nucleic Acids Res.* **11,** 1943–1954
38. Hamilton, R., Watanabe, C. K., and de Boer, H. A. (1987) *Nucleic Acids Res.* **15,** 3581–3593
39. Zaret, H. S., and Sherman, F. (1982) *Cell* **28,** 563–573
40. Mohana Rao, J. K., and Argos, P. (1986) *Biochim. Biophys. Acta* **869,** 197–214
41. Eisenberg, D., Schwarz, E., Komarony, M., and Wall, R. (1984) *J. Mol. Biol.* **179,** 125–142
42. Von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690
43. Early, A. E., Williams, J. G., Meyer, H. E., Por, S. B., Smith, E., Williams, K. L., and Gooley, A. A. (1988) *Mol. Cell. Biol.* **8,** 3458–3466
44. Broone, C., Sommer, S. S., Hensel, A., and Bussey, H. (1990) *J. Cell Biol.* **110,** 1833–1843
45. Lipke, P. N., Wojciechowicz, D., and Kurjan, J. (1989) *Mol. Cell. Biol.* **9,** 3155–3165
46. Li, W. B., Bzik, D. J., Horii, T., and Inselburg, J. (1989) *Mol. Biochem. Parasitol.* **33,** 13–26
47. Watzele, M., Klis, F., and Tanner, W. (1988) *EMBO J.* **7,** 1483–1488
48. Sharp, P. M., Tuohy, T. M. F., and Mosursky, K. R. (1986) *Nucleic Acids Res.* **14,** 5125–5143
49. Ferro-Novick, S., Hansen, W., Schauer, I., and Shekman, R. (1984) *J. Cell Biol.* **98,** 44–53
50. Haselbeck, A., and Tanner, W. (1983) *FEBS Lett.* **158,** 335–338
51. O'Hara, P. J., Horowitz, H., Eichinger, G., and Young, E. T. (1988) *Nucleic Acids Res.* **16,** 10153–10169
52. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
53. Hopp, T. P. (1986) *J. Immunol. Methods* **88,** 1–18