MOLECULAR AND CELLULAR BIOLOGY, June 1990, p. 3163–3173
0270-7306/90/063163-11$02.00/0

Vol. 10, No. 6

# Efficiency and Diversity of Protein Localization by Random Signal Sequences

CHRIS A. KAISER[1,2] AND DAVID BOTSTEIN[1,3]*

*Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139[1]; Department of Biochemistry, University of California, Berkeley, California 94720[2]; and Genentech, Inc., South San Francisco, California 94080[3]*

Received 27 November 1989/Accepted 26 March 1990

**Three randomly derived sequences that can substitute for the signal peptide of *Saccharomyces cerevisiae* invertase were tested for the efficiency with which they can translocate invertase or β-galactosidase into the endoplasmic reticulum. The rate of translocation, as measured by glycosylation, was estimated in pulse-chase experiments to be less than 6 min. When fused to β-galactosidase, these peptides, like the normal invertase signal sequence, direct the hybrid protein to a perinuclear region, consistent with localization to the endoplasmic reticulum. The diversity of function of random peptides was studied further by immunofluorescence localization of proteins fused to 28 random sequences: 4 directed the hybrid to the endoplasmic reticulum, 3 directed it to the mitochondria, and 1 directed it to the nucleus.**

Virtually all secreted proteins are synthesized with a hydrophobic amino-terminal signal peptide that is cleaved during the process of secretion. A large body of biochemical and genetic data indicates that the signal peptide is the primary determinant for segregation of proteins into the secretory pathway (5, 8, 11, 20, 32). Two complementary genetic approaches have been used to identify the elements within signal peptides that are required for function.

The first approach was used to identify point mutations in signal peptides that impair signal function. Mutations in secreted bacterial proteins that produce a complete block in translocation were isolated (5, 11). For example, in the signal sequence of the maltose-binding protein of *Escherichia coli*, there appear to be two amino acid positions that can be mutated to block maltose-binding protein export and give rise to a pronounced maltose growth defect (6). The absolutely defective mutations almost always introduce charged residues into the hydrophobic domain of the signal peptide, revealing the importance of hydrophobicity in signal peptide function. Other mutations do not produce an absolute block in translocation but reduce the translocation rate (4). These mutations are typically studied by kinetic experiments and define residues in the signal peptide that are required for optimal function (29).

The second approach was used to determine what kind of peptide sequences can functionally replace a known signal peptide. When the signal peptide of *Saccharomyces cerevisiae* invertase is replaced with essentially random peptides, about one-fifth of such peptides will direct invertase to the secretory pathway (16). Signal recognition appears to be based primarily, if not exclusively, on the hydrophobicity of a given peptide, since all random peptides with a domain of at least modest hydrophobicity function as translocation signals. This explains why few point mutations in signal sequences produce a frank secretory defect (1, 7, 21) and how the diverse naturally occurring signal peptides can all be recognized by a common secretory apparatus.

Because the assay used to identify functional substitutes for the invertase signal sequence is quite sensitive, peptides with a broad range of functional capacities were isolated. These sequences can be roughly divided into two phenotypic classes on the basis of the steady-state fraction of invertase that is glycosylated and therefore can be assumed to have been translocated into the lumen of the endoplasmic reticulum (ER). A functional class gives rise only to glycosylated invertase, and a partially functional class gives rise to both glycosylated and nonglycosylated invertase. Setting aside those sequences that show partial function, more than 10% of randomly isolated sequences exhibit in the steady state an efficiency of translocation that is indistinguishable from that of the wild-type invertase signal peptide (16). We were interested whether the function of these synthetic signal peptides could in some way be distinguished from that of the wild-type invertase signal peptide. One difference that has been noted is that most of the randomly isolated signal peptides are not cleaved and therefore must lack determinants required for recognition by signal peptidase (16). In this paper, we describe further tests of the function of three randomly isolated signal peptides and find that other than being defective for signal processing, they behave similarly to the native invertase signal peptide.

The first test addresses the possibility that as-yet-undefined information within mature invertase is necessary for the recognition of the synthetic signal peptides. If this is the case, most of the signal peptides identified in a collection of random sequences might be recognized as translocation signals only in the context of invertase. We tested three such peptides for the ability to function as secretion signals for *E. coli*, β-galactosidase, a protein that should not itself contain specialized secretory information. Sequences that function as secretion signals for invertase could also direct β-galactosidase to the ER, although at a somewhat lower efficiency than that of the native invertase signal sequence.

Another possible difference between the wild-type signal sequence and randomly isolated synthetic signal sequences is that the synthetic sequences may function at a greatly reduced efficiency. To test this, we determined the translocation rate governed by several different leader sequences and found both the wild-type *SUC2* signal sequence and the randomly isolated signal sequences to give rise to protein translocation in less than 6 min. Again, this shows that there is no pronounced difference between the function of the

* Corresponding author.

synthetic signal sequences and that of the native invertase signal sequence.

Finally, we investigated other possible fates of proteins with randomly isolated amino-terminal sequences by immunofluorescence microscopy. In addition to signals that direct proteins to the ER, both mitochondrial and nuclear localization signals were found within a collection of 28 essentially random peptide sequences. Taken together, these results imply that the cell discriminates between protein molecules destined for different cellular locations by several different recognition mechanisms, each able to recognize an extremely broad range of primary sequences.

## MATERIALS AND METHODS

**Strains, plasmids, and culture media.** *S. cerevisiae* DBY2449 is *MATα ura3-52 ade2-101 suc2-Δ9* (obtained from M. Carlson). *E. coli* DB6507 is HB101 (*leu pro thr hsdR hsdM recA*) with a Tn*5* insertion in the *pyrF* gene.

YEP liquid medium is 10 g of yeast extract (Difco Laboratories, Detroit, Mich.) per liter and 20 g of Bacto-Peptone (Difco) per liter. Solid medium contains 20 g of Bacto-Agar (Difco) per liter. SD-CAS liquid medium is 6.7 g of yeast nitrogen base without amino acids (Difco), 0.2% Bacto-Casamino Acids (vitamin free) (Difco) adsorbed with charcoal (Norite [fine]), and 2% glucose. Cell density was measured in a 1-cm quartz cuvette in a spectrophotometer (Lambda 3B; The Perkin-Elmer Corp., Norwalk, Conn.). One $A_{600}$ unit corresponds to approximately $10^7$ haploid yeast cells.

Plasmid pRB723 construction employed a DNA fragment carrying the *PGK1* gene promoter (14) with the base triplet ACC inserted directly upstream of the initiation codon to generate a *Nco*I site at this position (the generous gift of Don Moir at Collaborative Research, Waltham, Mass.). This fragment was ligated to the *Eco*RI site in the polylinker region at the amino terminus of the *lacZ* gene on plasmid pCGS63 (also obtained from Collaborative Research) after both ends were made blunt by filling in with DNA polymerase I, to produce pRB723. A *Bam*HI linker with the sequence CGCGGATCCGCG was inserted into the *Sma*I site of pRB576 (16), and the *PGK1* promoter fragment from pRB723 was inserted at this position to produce pRB722.

**Construction of library.** Human genomic DNA (60 μg) was digested with 100 U of each of the following restriction enzymes: *Alu*I, *Hae*III, *Rsa*I, and *Msp*I (New England BioLabs, Inc., Beverly, Mass.). The reaction was incubated at 37°C for 60 min in 0.5 ml of a solution consisting of 50 mM sodium chloride, 10 mM Tris hydrochloride (pH 7.5), and 10 mM magnesium chloride. DNA was precipitated in 0.3 M sodium acetate–60% (vol/vol) ethanol and then suspended in 0.15 ml of a solution consisting of 50 mM Tris hydrochloride (pH 7.5), 10 mM magnesium chloride, 0.1 mM dithiothreitol, 50 μg of bovine serum albumin per ml, and 40 mM each of the four deoxynucleoside triphosphates. To produce blunt DNA ends, 5 U of the large fragment of DNA polymerase I (New England BioLabs) was added and the reaction was incubated at 26°C for 30 min. The population of DNA fragments was fractionated on a 10% polyacrylamide gel, and fragments of 50 to 100 base pairs were isolated from the gel (18). Plasmid pRB722 DNA was digested with *Sma*I (New England BioLabs), treated with calf intestinal phosphatase (Boehringer Mannheim Biochemicals, Indianapolis, Ind.), phenol extracted to remove enzymes, and precipitated in ethanol. Linear plasmid DNA (5 μg) was ligated to the purified DNA fragments (0.5 μg) by incubation at 26°C for 18 h in ligase buffer (50 mM Tris hydrochloride [pH 7.5], 20 mM magnesium chloride, 20 mM dithiothreitol, 1 mM ATP). The ligated DNA was digested with *Sma*I to linearize vector DNA without an insert and then was used to transform *E. coli* DB6507. A total of 10 pools of about 1,000 transformants each were collected, and plasmid DNA was prepared separately from each pool. Plasmid DNA was then used to transform yeast strain DBY2449.

**Small-scale invertase assay.** Transformed yeast colonies were picked from selective plates, and approximately $10^6$ cells were suspended in 50 ml of lysis buffer (10 mM Tris hydrochloride [pH 7.5], 1% [vol/vol] Triton X-100, 0.1 mg of Zymolyase 100T [Seikagaku Kogyo Co., Tokyo, Japan] per ml) in the well of a multiwell microdilution dish. Cells became permeable after incubation at 26°C for 1 h. Substrate was added as 50 ml of a solution of 100 mM sucrose–60 mM sodium acetate (pH 5.2), and incubation was continued for 15 min. Sucrose hydrolysis was detected after the addition of 0.1 ml of assay cocktail (0.1 M Tris hydrochloride [pH 7.5], 0.1 mg of glucose oxidase [Boehringer Mannheim] per ml, 0.02 mg of peroxidase [Sigma Chemical Co., St. Louis, Mo.] per ml, 0.6 mg of *o*-dianisidine [Sigma] per ml) and incubation at 26°C for 30 min. The reaction was stopped by the addition of 0.1 ml of concentrated hydrochloric acid. Strains expressing cytoplasmic forms of invertase at only a few percent of the level of wild-type invertase are easily detected by this assay.

**Immunoblotting of invertase and β-galactosidase.** Yeast strain DBY2449 carrying different insert-containing derivatives of pRB722 and pRB723 was grown in SD-CAS with 0.1 mg of adenine per ml. Cell cultures (2 $A_{600}$ units) were washed once with 25 mM Tris hydrochloride (pH 7.5)–10 mM sodium azide, and cell pellets were lysed by agitation for 1 min on a vortex mixer in the presence of 20 ml of ESB (80 mM Tris hydrochloride [pH 6.8], 2% [wt/vol] sodium dodecyl sulfate [SDS], 0.01% bromphenol blue, 0.1 M dithiothreitol, 10% glycerol, 2 mM phenylmethylsulfonyl fluoride) and 0.1 g of glass beads and then heated to 95°C for 2 min. For endo-β-*N*-acetylglucosaminidase H (endo H) digestion, the extracts were diluted with 2 volumes of 30 mM sodium citrate (pH 5.2), endo H (25 U/mg) was added to a final concentration of 3 μg/ml, and the samples were incubated at 37°C overnight. One microliter of 1 M Tris hydrochloride (pH 6.8) and 6 μl of a solution of 30% (vol/vol) glycerol, 6% (wt/vol) SDS, and 0.3 M dithiothreitol were added to 18 ml of endo H-treated sample. The other samples were diluted with 3 volumes of ESB, and 15 μl of each was resolved on a SDS–7.5% polyacrylamide gel. Proteins were transferred electrophoretically to nitrocellulose filters, and invertase and β-galactosidase were identified by immunoblotting as described previously (14), except that a 1:50 dilution of mouse monoclonal anti-β-galactosidase antibody (tissue culture supernatant) and a 1:500 dilution of rabbit anti-mouse antiserum were used before $^{125}$I-labeled protein A (Amersham Corp., Arlington Heights, Ill.) was used to identify β-galactosidase antigen.

**Pulse labeling and immunoadsorption.** Cells were grown to exponential phase in SD medium supplemented with 2% glucose and 0.1 mg of adenine per ml. To induce invertase synthesis, cells were transferred to the same medium containing 0.1% glucose at a cell density of 1 $A_{600}$ unit per ml and aerated at 30°C for 1 h. Cultures (5 ml) were pulse labeled by the addition of 100 μCi of [$^{35}$S]methionine (Amersham), and after 6 min the chase was initiated by the addition of a 1/10 volume of 10 mg of methionine per ml (approximately 20% of the label is converted into a trichloro-

acetic acid-precipitable form in this time). Cells from 1-ml samples were pelleted by brief centrifugation (12,000 × *g*), suspended in 20 μl of a solution of 1% (wt/vol) SDS, 50 mM dithiothreitol, and 50 mM Tris hydrochloride (pH 7.5), and heated to 95% for 3 min. Samples were then lysed by agitation with 0.1 g of glass beads and diluted into 1 ml of a solution of 1% (vol/vol) Triton X-100, 50 mM Tris hydrochloride (pH 7.5), 0.1 M sodium chloride, and 20 μl of protein A conjugated to Sepharose CL-4B (50%, vol/vol) (Sigma). After 30 min of incubation at 4°C, cell debris and beads were cleared from the extract by centrifugation at 12,000 × *g* for 10 min. For precipitation of invertase, 1.5 μl of anti-invertase serum and an extract prepared from 10 $A_{600}$ units of cells with a deletion of the *SUC2* gene (yeast strain DBY2449) were added. For precipitation of β-galactosidase, 0.2 ml of mouse monoclonal anti-β-galactosidase (culture supernatant) was used. Immune complexes were allowed to form overnight at 4°C and were adsorbed with 20 μl of protein A-Sepharose CL-4B (invertase) or 20 μl of protein A-Sepharose CL-4B treated with 5 μl of rabbit anti-mouse antiserum (β-galactosidase). Immune complexes were washed four times with 1 ml of RIPA buffer (1% [vol/vol] Triton X-100, 1% [wt/vol] deoxycholic acid, 0.1% [wt/vol] SDS, 150 mM sodium chloride, 10 mM Tris hydrochloride [pH 7.5]). The final pellet was suspended in 40 μl of loading buffer and heated to 95°C for 3 min. Samples (20 μl) were resolved by SDS-polyacrylamide gel electrophoresis on 7.5% gels treated with Autofluor and exposed to XAR-5 film (Eastman Kodak Co., Rochester, N.Y.) at −70°C.

**DNA sequencing.** DNA fragments containing the human genomic inserts were isolated from the derivatives of pRB722 and were inserted into the M13-based sequencing vectors Tg130 and Tg131 (23) (Amersham). For all the inserts, both DNA strands were sequenced independently by the method of Sanger et al. (27) by using T7 DNA polymerase in the Sequenase kit (U.S. Biochemical Corp., Cleveland, Ohio) and [$^{35}$S]dATP (Amersham).

**Immunofluorescence.** Cells were grown at 30°C in SD-CAS with 2% glucose. Cells were rapidly fixed by filtering the culture through a 0.45-mm-pore-size filter (Millipore Corp., Bedford, Mass.), suspending the cells in 5 ml of 50 mM sodium phosphate (pH 6.5)–4% (vol/vol) formaldehyde (Mallinckrodt, Inc., St. Louis, Mo.), and allowing the suspension to stand at 25°C for 1 h. Cells were washed twice by centrifugation and suspension in a solution of 1.4 M sorbitol and 50 mM Tris hydrochloride (pH 7.5). Cell walls were disrupted by incubation for 30 min at 30°C in 2 ml of a solution of 1.4 M sorbitol, 50 mM Tris hydrochloride (pH 7.5), 30 mM 2-mercaptoethanol, and 50 μg of Zymolyase 100T. Spheroplasts were centrifuged at 2,000 × *g* for 5 min and suspended in 1 ml of 1.4 M sorbitol–50 mM Tris hydrochloride (pH 7.5). Multiwell glass slides were treated by applying 15 μl of 1 mg of poly-L-lysine per ml to each well for 10 min, submerging the slide in distilled water for 10 min, and allowing the slide to dry completely. The spheroplast suspension was applied to the slide (15 μl per well) and allowed to stand for about 5 min. Excess liquid was removed by aspiration, and primary antibody was applied in TSA buffer (100 mM sodium chloride, 20 mM Tris hydrochloride [pH 7.5], 100 μg of bovine serum albumin per ml) and allowed to incubate at 25°C for 45 min. Either monoclonal anti-β-galactosidase in tissue culture supernatant (a gift of J. Teem) or affinity-purified rabbit anti-invertase antibody (a gift of S. Ferro-Novick and H. Ruohola) was used at a 1:50 dilution. Cells were washed three times with TSA buffer and then incubated in 15 ml of secondary antibody in TSA for 45 min at 25°C. Secondary antibody was either fluorescein-conjugated goat anti-rabbit immunoglobulin G (invertase staining) at a 1:500 dilution or rabbit anti-mouse immunoglobulin G at a 1:500 dilution (a gift of T. Briner) followed by a 1:500 dilution of fluorescein-conjugated goat anti-rabbit antibody (Organon Teknika, Malvern, Pa.) (β-galactosidase staining). Cells were stained with 0.5 mg of 4,6-diamidino-2-phenylindole (DAPI) per ml in TSA for 5 min. Cells were washed three times with TSA by aspiration and mounted in 10 ml of a solution of 90% glycerol, 1 mg of *p*-phenylenediamine (pH 9) per ml, 0.85% (wt/vol) sodium chloride, and 66 mM sodium phosphate. Micrographs were made with a Zeiss Standard photomicroscope by using either epifluorescence or Nomarski optics with a Plan-Neofluor objective (63×; NA, 1.25). Hypersensitized Kodak Technical Pan 2415 film (Lumicon, Livermore, Calif.) was used. In some cases, the intensity of staining of individual cells was variable. Individual cells do not appear to differ in cell permeability to the antibody molecules, because identical cell preparations showed uniform staining of cells when labeled with antibody specific for alcohol dehydrogenase I, a constitutive cytoplasmic protein (data not shown). Therefore, the variability in staining is probably a result of differences in the copy number of the plasmids which use the 2μm plasmid origin of replication.

## RESULTS

**Construction of peptide-invertase fusions.** Plasmid pRB722 was designed to analyze the functions of different leader peptides fused to invertase. This plasmid contains the promoter sequences, including the initiation codon for the 3-phosphoglycerate kinase gene (*PGK1*) of *S. cerevisiae*, fused to the coding sequence of mature invertase lacking the signal peptide. At the fusion junction, there is a DNA linker which contains four restriction enzymes sites (Fig. 1). A DNA fragment inserted into the *Sma*I site in the linker will be transcribed and translated, since this site is on the 3′ side of the initiation codon. If the inserted sequence is of the correct length to join the *PGK1* initiation codon in frame with the invertase sequence and if the insert does not contain any termination codons, then this hybrid gene will specify a protein composed of an amino-terminal peptide encoded by the inserted fragment fused to mature invertase. A given leader peptide-coding sequence can then be separated from the *SUC2*-coding sequence and fused to β-galactosidase by transferring the coding sequence for the leader peptide to plasmid pRB723 (Fig. 1). Both pRB722 and pRB723 carry selectable markers (*bla* and *URA3*) and replication sequences (ColE1 *ori* and *S. cerevisiae ARS1* and *CEN4* on pRB722 and 2μm *ori* on pRB723) that allow the plasmids to be maintained in both *S. cerevisiae* and *E. coli*.

**Isolation of peptide-invertase fusions.** A diverse collection of leader sequences fused to invertase was isolated from a library of short fragments of human genomic DNA inserted into the *Sma*I site of pRB722 (see Materials and Methods). Plasmid library DNA was recovered from $10^4$ *E. coli* transformants and introduced by transformation into *S. cerevisiae* DBY2449 that has a deletion of the chromosomal *SUC2* gene.

It was necessary to screen many transformants to find rare plasmids that encode an upstream sequence fused to invertase that is in the correct reading frame and that does not contain a termination codon. The most general screen for the presence of peptide-invertase hybrid proteins would be to test individual yeast transformants for invertase antigen by

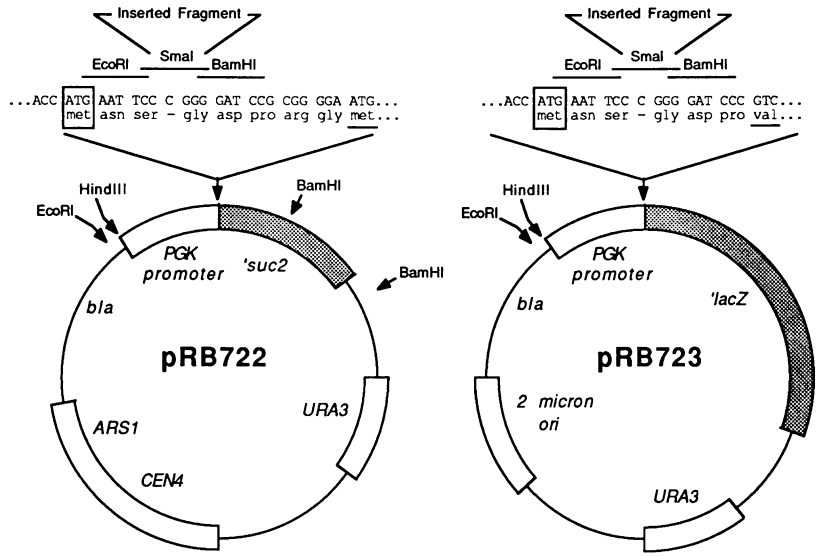


FIG. 1. Plasmid vectors used to generate fusions between random peptide-coding sequences and the *SUC2* gene (pRB722) or the *lacZ* gene (pRB723). For each plasmid, the linker joining the *PGK1* promoter and coding sequences is shown. The boxed codons correspond to the *PGK1* initiation codon. The underlined codon in pRB722 is the initiation codon of the normal cytoplasmic form of invertase. The underlined codon in pRB723 corresponds to the amino acid 9 of β-galactosidase. The ACC triplet that is upstream of the initiation codon is not present in the wild-type *PGK1* promoter.

immunoblotting cell extracts resolved by SDS-gel electrophoresis. In a previous experiment, we screened 431 transformants by immunoblotting and found 16 that produced invertase with an additional leader peptide (16). All these transformants also produced invertase enzymatic activity, indicating that most if not all fusion proteins can be identified by enzymatic activity alone. An activity screen was used to detect clones that produce invertase, since it was much easier to perform than a gel assay.

Transformed colonies were lysed by treatment with Zymolyase and nonionic detergent to allow detection of intracellular enzyme, and the lysates were assayed for invertase activity. Conveniently, the *PGK1* initiation codon of pRB722 is not in frame with the *SUC2*-coding sequence, so that the vector alone without an insert would not produce detectable activity in this screen. Microdilution dishes were used to screen 841 transformants, and as expected, about 5% (40 of 841) of the transformants produced active enzyme. The form of invertase protein produced by each of these transformants was examined by resolving cell extracts on nondenaturing polyacrylamide gels and staining for invertase activity (data not shown). Of the 40 transformants that expressed invertase, 10 produced a form of the enzyme that is indistinguishable by gel assay from the normal cytoplasmic form of invertase. The inserted DNA fragment in one of these clones (1E9) was sequenced and found to have a termination codon in the same reading frame as that of the *PGK1* initiation codon. In this case, the cytoplasmic form of invertase was probably produced as a result of an inefficient translational initiation at the second AUG codon on the mRNA. This transformant class was not studied further. The other transformants all produced forms of invertase with lower mobilities in nondenaturing gels than those of the normal cytoplasmic form of invertase. These were likely to be invertase molecules fused to additional peptide sequences. Two of the transformants in this class produced very low levels of invertase and also were not studied further. Plasmid DNA was isolated from each of the remaining 28 clones, amplified in *E. coli*, and returned to DBY2449 by transformation. Each of these transformants was then assayed for invertase that had entered the secretory pathway.

**Glycosylation of peptide-invertase fusions.** On entering the ER, secreted invertase is modified by the addition of approximately 10 N-linked oligosaccharide chains (30). This modification increases the apparent molecular weight of invertase from 55 kilodaltons to about 90 kilodaltons (13). On transport to the Golgi apparatus, invertase is further modified by the addition of the high-mannose outer chains causing a variable increase in the apparent molecular weight to about 120 kilodaltons. These different forms of invertase are easily distinguished on SDS gels. Extracts from strains with different insert-bearing plasmids were resolved by SDS-polyacrylamide gel electrophoresis, and invertase antigen was detected by immunoblotting with anti-invertase serum and $^{125}$I-labeled protein A. Of the 28 leader peptides, 3 (4H11, 6C6, and 8B5) clearly gave rise to glycosylated invertase (Fig. 2). Four others (1B9, 2C3, 6G8, and 1G10) led to the production of less-distinct high-molecular-weight forms that were also likely to be glycosylated invertase, although it was difficult in these cases to distinguish glycosylated invertase from background in the gel lane. All the other strains gave rise to forms of invertase with mobilities slightly lower than that of the normal cytoplasmic form of invertase. Presumably, these are invertase molecules with amino-terminal leader sequences that do not function as secretion signals and thus do not lead to invertase glycosylation.

**Secretion of peptide-invertase fusions.** The fraction of invertase activity that is at the cell surface was determined for each strain in an independent test for invertase secretion. To obtain the invertase levels outside and inside the plasma membrane, intact cells and lysed spheroplasts were assayed independently for invertase activity. Previously, we found that strains expressing invertase without a functional signal sequence produce less than 2% of the total activity at the cell surface. In contrast, between 10 and 90% of the activity is at the surface for forms of the enzyme with synthetic signal sequences that express different degrees of function (16). Extracellular activity is not an absolute criterion for signal

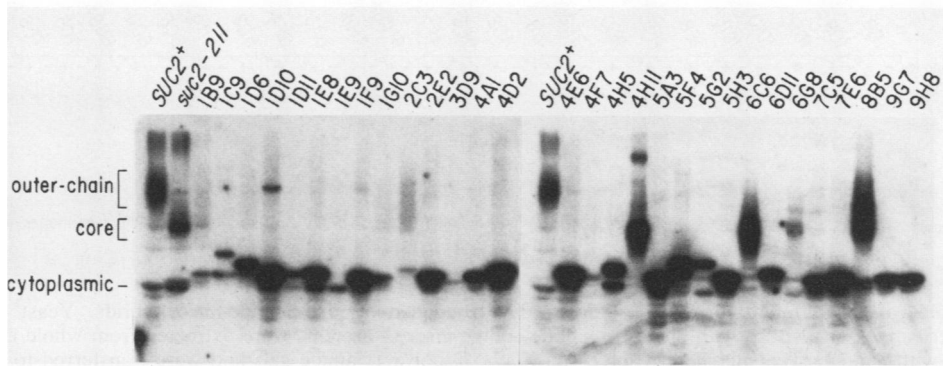


FIG. 2. Immunoblot of the forms of invertase produced by the different peptide-invertase hybrids. Yeast strain DBY2449 with a deletion of the *SUC2* gene was transformed with derivatives of pRB722 carrying different peptide-coding inserts. Protein extracts were prepared from whole cells and were resolved on SDS–7.5% polyacrylamide gels. Proteins were transferred to nitrocellulose sheets, and invertase was identified with anti-invertase antibodies and $^{125}$I-labeled protein A. The wild-type *SUC2* gene on plasmid pRB420 (14) produced both nonglycosylated cytoplasmic invertase and outer chain-glycosylated invertase (lanes SUC2$^+$). The *suc2-211* allele is used as a marker for core-glycosylated invertase (15).

translocation function, since some leader peptides led to high levels of invertase glycosylation, but the extracellular enzyme levels were low because invertase was inefficiently transported to the cell surface after entering the ER (16). The seven strains for which glycosylated invertase was detected all express more than 10% of the invertase activity at the cell surface (Table 1). In addition, peptide 4F7 gave rise to a significant fraction of the total invertase produced as extracellular enzyme, although no glycosylated invertase was detected on the immunoblot for this hybrid protein. The total amount of invertase produced by 4F7 is small (Table 1), and it is possible that the band of glycosylated invertase is too faint to be seen on the immunoblot.

Since there is no obvious absolute criterion for defining signal sequence function, the distinction that we make between functional and nonfunctional leader sequences must be somewhat arbitrary. We will define as a minimally functional leader sequence one that gives rise to either glycosylated invertase of an immunoblot or a significant fraction of the invertase activity on the cell surface. By these criteria, the sequences 4H11, 6C6, 8B5, 1B9, 6G8, 2C3, 1G10, and 4F7 are all functional, although their efficiencies and absolute levels of expression differ widely. Thus, within the sample of 28 sequences, the fraction that is at least minimally functional (8 of 28) is consistent with our expectation on the basis of previous experiments in which about one-fifth of leader peptides encoded by fragments of human DNA were found to have at least a partial function as secretion signals (16). Of these sequences, 4H11, 6C6, and 8B5, or 3 of 28, appear to be efficient secretion signals and give rise to essentially 100% translocation of invertase.

**Glycosylation of peptide–β-galactosidase hybrids.** Each of the leader peptides was also tested for the ability to direct β-galactosidase to the secretory pathway. We found that the fusion of the first 19 amino acids of preinvertase to β-galactosidase will cause at least part of the hybrid protein to be translocated into the lumen of the ER, since β-galactosidase itself can serve as a substrate for N-linked carbohydrate addition and since two or three of the six potential carbohydrate acceptor sequences (Asn-X-Ser/Thr) in β-galactosidase are modified when the *SUC2* signal sequence is present (see Fig. 5). All the peptides were tested for the ability to act as a secretion signal for β-galactosidase by using glycosylation as a criterion for entry into the secretory pathway.

The leader peptide-coding sequences with the *PGK1* pro-

TABLE 1. Properties of leader peptides fused to invertase or β-galactosidase

| Leader peptide[a] | Invertase | | | | β-Galactosidase | |
|---|---|---|---|---|---|---|
| | Total activity (units)[b] | Surface activity (%)[c] | Glycosylation[d] | Location (immunomicroscopy)[e] | Glycosylation | Location (immunomicroscopy) |
| 4H11 | 870 | 11 | + | Perinuclear | + | Perinuclear |
| 6C6 | 900 | 26 | + | Perinuclear | + | Perinuclear |
| 8B5 | 1,220 | 53 | + | Mitochondrial | + | Perinuclear |
| 6G8 | 150 | 13 | ± | Perinuclear[f] | + | Perinuclear |
| 2C3 | 150 | 64 | ± | Mitochondrial | NT | NT |
| 1B9 | 70 | 36 | ± | Cytoplasmic[g] | − | Cytoplasmic[g] |
| 1G10 | 54 | 41 | ± | Cytoplasmic[g] | − | Cytoplasmic[g] |
| 4F7 | 38 | 24 | − | Cytoplasmic[g] | − | Cytoplasmic[g] |
| 1C9 | 42 | 4 | − | Mitochondrial | − | Cytoplasmic |
| 5G2 | 170 | 3 | − | Nuclear | − | Nuclear |
| 1D6 | 170 | 6 | − | Cytoplasmic | − | Cytoplasmic |
| 1D10 | 1,140 | 1 | − | Cytoplasmic | − | Cytoplasmic |
| 1D11 | 46 | 4 | − | Cytoplasmic | − | Cytoplasmic |
| 1E8 | 460 | 2 | − | Cytoplasmic | − | Cytoplasmic |
| 1F9 | 860 | 2 | − | Cytoplasmic | − | Cytoplasmic |
| 2E2 | 210 | 3 | − | Cytoplasmic | − | Cytoplasmic |
| 4A1 | 110 | 2 | − | Cytoplasmic | − | Cytoplasmic |
| 4D2 | 730 | 1 | − | Cytoplasmic | − | Cytoplasmic |
| 4E6 | 820 | 1 | − | Cytoplasmic | − | Cytoplasmic |
| 4H5 | 950 | 1 | − | Cytoplasmic | − | Cytoplasmic |
| 5A3 | 1,010 | 1 | − | Cytoplasmic | − | Cytoplasmic |
| 5F4 | 850 | 2 | − | Cytoplasmic | − | Cytoplasmic |
| 5H3 | 670 | 1 | − | Cytoplasmic | − | Cytoplasmic |
| 6D11 | 620 | 1 | − | Cytoplasmic | − | Cytoplasmic[h] |
| 7C5 | 330 | 2 | − | Cytoplasmic | − | Cytoplasmic |
| 7E6 | 680 | 3 | − | Cytoplasmic | − | Cytoplasmic |
| 9G7 | 260 | 2 | − | Mitochondrial | − | Cytoplasmic |
| 9H8 | 140 | 2 | − | Cytoplasmic | − | Cytoplasmic |

[a] Strain DBY2449 (*MATα ura3-52 ade2-101 suc2-Δ9*) was transformed with either plasmid pRB722 (*SUC2* fusion) or pRB723 (*lacZ* fusion) carrying the indicated leader peptide.

[b] Total activity is the sum of the assays of invertase activities of intact cells and lysed spheroplasts (15).

[c] Fraction of total activity that was assayed on intact cells.

[d] N-linked glycosylation of invertase and β-galactosidase was detected by mobility shifts seen with SDS-polyacrylamide gel electrophoresis.

[e] The predominant pattern of cellular localization seen by indirect immunofluorescence microscopy of fixed cells.

[f] Perinuclear staining was relatively faint in this case.

[g] Faint nuclear staining was also seen in these strains.

[h] Some small vesicles were stained in this strain.

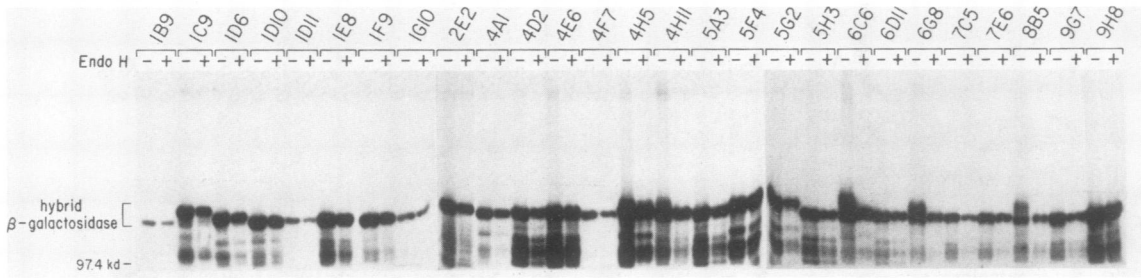


FIG. 3. Immunoblot of the different forms of β-galactosidase produced by the peptide-*lacZ* hybrids. Yeast strain DBY2449 was transformed with derivatives of pRB723 with different peptide-coding inserts. Proteins were extracted from whole cells, and half of each sample was treated with endo H. Proteins were resolved on SDS–5% polyacrylamide gels and were transferred to nitrocellulose sheets. β-Galactosidase was identified with mouse monoclonal anti-β-galactosidase antibodies, rabbit anti-mouse antibodies, and $^{125}$I-labeled protein A. The position of the predominant species related to β-galactosidase is indicated. The lower-molecular-weight forms are probably degradation products of β-galactosidase. kd, Kilodaltons.

moter were isolated from the plasmid derivatives of pRB722 on *Hin*dIII-*Bam*HI fragments and were fused to the *lacZ* gene on plasmid pRB723. Plasmids pRB722 and pRB723 were designed so that this subcloning manipulation conserves the reading frame and results in gene fusions which produce hybrid β-galactosidase molecules with leader peptides fused to the amino terminus. Peptide 2C3 was not subcloned, because it contains a *Hin*dIII site within the inserted sequence. As with the invertase hybrid proteins, glycosylation was detected by using a gel assay. Cell extracts were resolved by SDS-polyacrylamide gel electrophoresis, and the hybrid β-galactosidase molecules were identified by immunoblotting with anti-β-galactosidase antibodies. To facilitate detection of the small fractional change in mass that results from glycosylation, extracts treated with endo H to produce unglycosylated proteins were used for comparison.

Peptides 4H11, 6C6, 8B5, and 6G8 all gave rise to some β-galactosidase glycosylation (Fig. 3). Therefore, the leader peptides that functioned as efficient secretion signals for invertase also guided β-galactosidase to the ER. All the other leader peptides, including those that acted as inefficient secretion signals for invertase, had no detectable activities as β-galactosidase secretion signals. All the samples also produced immunoreactive molecules of lower molecular weight than intact β-galactosidase that are probably proteolytic degradation products of peptide–β-galactosidase fusion proteins.

**Rate of glycosylation of invertase hybrids.** Wild-type invertase is rapidly translocated into the ER and glycosylated; after cells expressing invertase are pulse labeled for 5 min, all the labeled invertase molecules are in a glycosylated state (28). If the synthetic signal peptides isolated here act on the same pathway as that of the wild-type signal peptide, then they too should promote rapid translocation of invertase. We examined the kinetics of glycosylation of the three peptide-invertase hybrids (4H11, 6C6, and 8B5) that produced abundant glycosylated forms of invertase. Cells expressing these hybrid proteins were pulse labeled with [$^{35}$S]methionine for 6 min, and the chase period was initiated by the addition of an excess of unlabeled methionine. Cell growth was stopped, and proteins were extracted rapidly by boiling the cells in 1% SDS and 50 mM dithiothreitol. Peptide-invertase hybrid proteins were recovered by immunoadsorption with anti-invertase antibodies. For wild-type invertase and the three peptide-invertase hybrids, most of the protein is in a core-glycosylated state after pulse labeling (Fig. 4). These three synthetic signal peptides cause invertase to be translocated and glycosylated in less than 6 min.

The hybrids with peptides 4H11 and 6C6 persist in a core-glycosylated state, indicating that these proteins remain in the ER throughout the 30-min chase period, presumably because the signal peptide is not cleaved. The hybrid with peptide 8B5 appears to be transported slowly from the ER; it is converted to an outer chain-glycosylated state with a half time of about 30 min, compared with the wild-type, which is converted to the fully glycosylated from with a half time of less than 5 min. As expected, the wild-type *SUC2* gene produced a nonglycosylated form of invertase which is stable throughout the 30-min chase period. Both hybrids with peptides 6C6 and 8B5 produced nonglycosylated forms

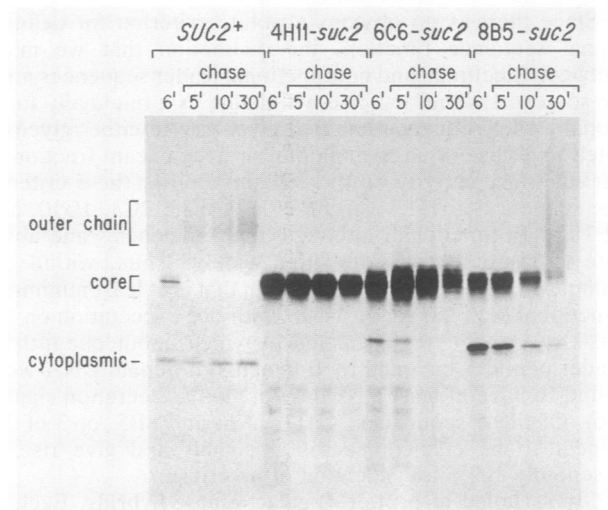


FIG. 4. Kinetics of glycosylation of the peptide-invertase hybrids. Cells containing derivatives of pRB722 containing different leader peptide-coding inserts were pulse labeled with [$^{35}$S]methionine for 6 min. The chase period was initiated by the addition of excess unlabeled methionine. Protein extracts were prepared by rapid boiling in the presence of SDS and dithiothreitol, and forms of invertase were isolated by immunoadsorption with anti-invertase antibodies. Proteins were resolved on SDS–7.5% polyacrylamide gels that were subjected to autoradiography. The positions of the core-glycosylated, outer chain-glycosylated, and nonglycosylated cytoplasmic forms of invertase are indicated. A difference in the strength of the *SUC2* and the *PGK1* promoters probably accounts for the different total invertase levels produced by the wild-type and fusion genes.

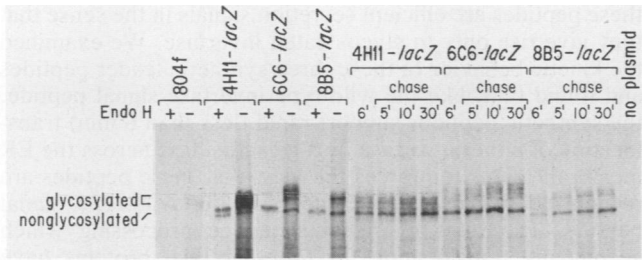


FIG. 5. Kinetics of glycosylation of peptide–β-galactosidase hybrids. Cells containing derivatives of pRB723 with different peptide-coding inserts were pulse labeled with [$^{35}$S]methionine. The chase period was initiated by the addition of an excess of unlabeled methionine. Proteins were extracted from cells, and forms of β-galactosidase were isolated by immunoadsorption with monoclonal anti-β-galactosidase antibodies. One sample from each strain was treated with endo H. Proteins were resolved on SDS–5% polyacrylamide gels and were subjected to autoradiography. The sample labeled 804f is a fusion protein containing the 19-amino-acid signal sequence of invertase fused to β-galactosidase.

as well. However, both of these proteins were unstable and had half lives of about 5 min. Since these are minor species, it is not possible to tell whether their disappearance is due to degradation or conversion to a glycosylated state.

**Rate of glycosylation of β-galactosidase hybrids.** The kinetics of glycosylation of the same leader peptides fused to β-galactosidase was also examined. As with invertase, the β-galactosidase hybrids were glycosylated within a 6-min labeling period. The glycosylated β-galactosidase appeared as a ladder of bands of greater relative molecular mass than that of the form of β-galactosidase in the endo H-treated samples (Fig. 5). For all three hybrid proteins, only about half the molecules were glycosylated and the relative proportion of glycosylated to nonglycosylated molecules remained constant over a 30-min chase period (Fig. 5). There are two populations of molecules: those that are translocated and glycosylated rapidly and those that are never glycosylated. The unglycosylated class of molecules either has failed to enter the ER or has entered the ER but is in some way unsuitable as substrates for glycosylation.

**Primary structure of the leader peptides.** The sequence of the DNA encoding each of the synthetic leader peptides was determined. All but one of the sequences had an open reading frame that began with the initiation codon for the *PGK1* gene and was in frame with the *SUC2*-coding sequences. The derived amino acid sequences are shown in Fig. 6. For the exceptional sequence (6G8), the *PGK1* initiation codon and the *SUC2*-coding sequence were not in the same reading frame. Either there was a mistake in sequence determination or there is a high frequency of translational frameshifting when this sequence is expressed. All the inserted sequences were found to be different from one another, except for sequences 1B9 and 1G10. Since these sequences were isolated from the same pool of *E. coli* transformants, they are almost certainly derived from the same recombinant molecule.

Previously, we found that functional and nonfunctional synthetic leader peptides can be separated on the basis of a combination of hydrophobicity and charge density (16). The sequences isolated here were analyzed in the same way by plotting for each sequence the hydrophobicity of the most hydrophobic segment of 10 amino acids against the fraction of total amino acids in each sequence that are charged. The sequences that were at least minimally functional as secre-

1B9 MNSPVCSYIVLILQVKLQCYLKGDPRGM...
1C9 MNSPSCPVQEGGGGVSPPPGSHLDCGRPILCKLVKFTIIKCYEELRELALSQKTKQQQQQQKTIKLC LLDRRGPWCHPLGGLREMSREKGLGAAWALLRQGCSGDTIMCVKGLCNQGDPRGM...
1D6 MNSLLCADLQLPHEVEPLSLQLTLCGPLWRTASLQPQKEEKETGMTGMSHRSRPLRFLFPFYSGDPK GDPRGM...
1D10 MNSPCCLYDLAKSSAFPKPQGDPRGM...
1D11 MNSPVSSHLLLNRFQIIFFLNKHRGDPRGM...
1E8 MNSPYLRVESRGLEEEQNKKNYLSGGDPRGM...
1F9 MNSPNFCICSRDGLSPCQGDPRGM...
1G10 MNSPVCSYIVLILQVKLQCYLKGDPRGM...
2C3 MNSPPPPRFTPFSCLSLPSRHISRGCWENQVNTNKSQSLAHGPSFFFFFFFLRQSLALLHK GDPRGM...
2E2 MNSPRLECSGAISALCNLRLPGDPRGM...
4A1 MNSPTRSGEDPFGFPSLSPEGDPRGM...
4D2 MNSPKNLEKRLDEWLTRINSITTTLNDPMEGDPRGM...
4E6 MNSPAQKLSPASTGTAKAAGKVILEGDPRGM...
4F7 MNSPGITGTHHYAWLIFCIFSRDGVSSCWGDPRGM...
4H5 MNSPKIKTKEEKHESGCDLPSLPKSISPPDHSSLTREEAERKLYSGAQTDGKFLYVGGDPRGM...
4H11 MNSPKKDIHTPPLKYPGGQHRGASCCFVFIMCLLYRICGICGDPRGM...
5A3 MNSPAEQGDKKKPQIHFSMEGDPRGM...
5F4 MNSPYCNAKDVPLCPELVISAPVPLEWLPWTPAETPVSRARNHSEYLCGDRGQTNPDKLKSHAEVHP KTEGGSKEAEKKQGDPRGM...
5G2 MNSPCSEPRSRHCTPAWATERDSVSKKKKKKKSRLEERWSKEGEKPAGFFSPKPLLAWTGSPIQSNW ARERNKGHPNWKRGSQTSGDPRGM...
5H3 MNSPPHQGLFMRRHPTAGAPGRILPGAQHQGDPRGM...
6C6 MNSPMISLPTTQKCANTTGHSSNRFQLFFGFSSLKTLFLSLLLMEIWEGDPRGM...
6D11 MNSPSRSQRPSATTQVFSLGQAIKFDVFTQYTSLIYSHNKMLGKTQGDPRGM...
7C5 MNSPHWLCHVYMRNERKQALSEERGDPRGM...
7E6 MNSPPSVNAVSMFPSIPHLEVFLPRGPGPTQGDPRGM...
8B5 MNSPPCTCLSSMLITGFIFLSSLQGDPRGM...
9G7 MNSPLFKTARSLLLLRENKLIRNIWEGDPRGM...
9H8 MNSPRVSSEGSTGEASTSKGDPRGM...

FIG. 6. Primary structures of the different leader peptides. The DNA sequence for each peptide-coding insert in plasmid pRB722 was determined, and the inferred amino acid sequence is shown. Each sequence begins with the methionine codon derived from the *PGK1* gene and ends with the methionine codon corresponding to the initiation codon for the normal cytoplasmic form of invertase. A single open reading frame was not found for the 6G8 sequence, and therefore it is not shown. Note that the sequences for 1B9 and 1G10 are identical. The single-letter amino acid code is as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

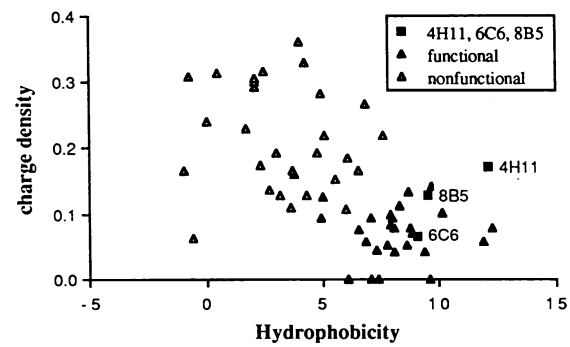


FIG. 7. Hydrophobicity of the leader peptides. For each peptide, the hydrophobicity of the most hydrophobic stretch of 10 residues is plotted against the fraction of residues that are charged (Lys, Arg, Glu, and Asp), as previously described (16). Peptides isolated in this study were combined with previously isolated peptides (16).

tion signals for invertase were separated from the nonfunctional sequences on the basis of these parameters (Fig. 7).

**Subcellular localization of hybrid proteins.** As another test of the ability of the leader peptides to act as localization signals, we located the hybrid invertase and β-galactosidase molecules within cells by indirect immunofluorescence microscopy. Cells expressing peptide–β-galactosidase hybrids were fixed with formaldehyde, converted to spheroplasts, and treated with mouse monoclonal anti-β-galactosidase antibody followed by fluorescein isothiocyanate-conjugated rabbit anti-mouse immunoglobulin G. Three general localization patterns were found. Most of the strains had diffuse staining throughout the cell body, suggesting that the β-galactosidase was located predominantly, if not exclusively, in the cytoplasm (Table 1) (Fig. 8i and j). A perinuclear staining pattern for the four strains that gave rise to glycosylated hybrid β-galactosidase molecules encoded by fusion to peptides 4H11, 6C6, 8B5, and 6G8 (Table 1) (Fig. 8c and d) was seen. This pattern of localization was identical to that of β-galactosidase fused to the wild-type *SUC2* signal peptide (data not shown), supporting the idea that peptides 4H11, 6C6, 8B5, and 6G8 act to direct β-galactosidase to the same cellular location as that specified by the wild-type invertase signal peptide. Staining coincident with the DAPI-stained nucleus was seen for peptide 5G2 fused to β-galactosidase (Fig. 8g and h). Therefore, peptide 5G2 acts as a nuclear localization signal for β-galactosidase.

Peptide-invertase hybrids stained with anti-invertase antibodies showed four different patterns of subcellular localization. As with the β-galactosidase hybrids, most of the invertase hybrid proteins had staining of the whole cell body, indicating a cytoplasmic location (Table 1). The three peptide-invertase fusions that produced mostly core-glycosylated invertase (4H11, 6C6, and 6G8) showed a perinuclear staining pattern. Thus, as with the β-galactosidase fusion proteins, perinuclear localization was seen for translocated proteins that were retained in the ER. The peptide 8B5-invertase hybrid which was efficiently glycosylated did not show perinuclear staining. This protein was found with predominantly outer-chain glycosylation in the steady state and therefore may not have accumulated in the ER to levels sufficient to give rise to visible perinuclear staining.

The peptide 5G2-invertase hybrid was located in the nucleus (Fig. 8e and f). Therefore, this peptide serves as a nuclear localization signal for invertase as well as β-galactosidase.

The hybrid invertase proteins produced by four of the hybrid genes (8B5, 1C9, 2C3, and 9G7) appeared to be located in the mitochondria. Within these cells, discrete elongated bodies were stained. These were usually at the periphery of the cytoplasm and were always coincident with DAPI-stained mitochondrial DNA (Fig. 8k and l). These sequences appear to be mitochondrial localization signals for invertase but not β-galactosidase. Two of these peptides (2C3 and 8B5) also gave rise to glycosylated invertase and therefore function as both mitochondrial and ER localization signals.

## DISCUSSION

A remarkably large fraction of essentially random peptide sequences can replace the normal invertase signal peptide as a secretion signal. A collection of 28 peptide-coding sequences were isolated from fragments of human genomic DNA by their ability to be translated in yeast cells when fused to the amino terminus of the *SUC2* gene. Three of these peptides are efficient secretion signals in the sense that they give rise only to glycosylated invertase. We examined the kinetic behavior of these three synthetic leader peptides and found that, like the wild-type invertase signal peptide, the synthetic peptides support rapid (less than 6 min) translocation of either invertase or β-galactosidase across the ER membrane. This reinforces the view that these peptides are recognized in the same manner as the wild-type signal peptide. The rates of signal sequence processing which accompany translocation for four bacterial proteins have been analyzed by pulse chase, and in all cases, signal processing is completed in less than 30 s (19, 25, 29). Therefore, it is possible that shorter pulse labelings could reveal distinctions between the native and randomly isolated signal peptides.

When each of three synthetic signal peptides was fused to the amino terminus of β-galactosidase, the hybrid proteins were partially glycosylated on β-galactosidase sequences and therefore had entered the ER. Assuming that β-galactosidase carries no localization information itself, the fused peptide sequences must themselves be recognized by the secretory apparatus.

In general, β-galactosidase appears to be translocated at a lower efficiency than invertase. For example, peptide 4H11 caused complete core glycosylation of invertase, whereas only about half the β-galactosidase molecules were glycosylated when fused to this peptide, and peptide 1B9 caused some invertase glycosylation when fused to invertase but did not give rise to any detectable β-galactosidase glycosylation. The reduced efficiency of β-galactosidase translocation does not appear simply to be a result of saturation of the translocation apparatus by overproduced β-galactosidase fusion proteins, since immunoprecipitation experiments carried out in antibody excess indicate that the β-galactosidase fusions are expressed at a lower level than the corresponding invertase fusion proteins. Therefore, some property of mature invertase must facilitate either signal recognition or translocation across membranes. This could be a sequence within invertase that increases the efficiency of binding of a leader peptide to a signal recognition element. Alternatively, invertase but not β-galactosidase may contain sequences that prevent dissociation of the protein from the membrane once translocation has begun.

The behavior of these synthetic signal peptides sheds light on the question of how secretory proteins evolve. It is likely that secretory proteins typically evolve from cytoplasmic proteins. Given that one-fifth of random sequences are recognized to some degree as signal peptides, it is conceivable that a cytoplasmic protein could acquire a signal peptide by a single mutational event. A DNA rearrangement that fuses additional coding sequences to the amino terminus of a cytoplasmic protein should frequently produce a protein with the capacity to enter the secretory pathway. Such mutations have been isolated as intragenic suppressors of certain secretion-defective *SUC2* mutations (24). Our observation here that about 10% of randomly isolated peptides fused to β-galactosidase cause at least a portion of the protein to be translocated into the ER demonstrates the feasibility of this mechanism for the conversion of a cytoplasmic protein into a secretory protein.

The cellular location of the hybrid proteins as determined by immunofluorescence microscopy provides an independent assessment of the ability of these peptide sequences to act as localization signals. The invertase and β-galactosidase fusion proteins that are predominantly core glycosylated in the steady state (4H11, 6C6, and 6G8) show localization to a

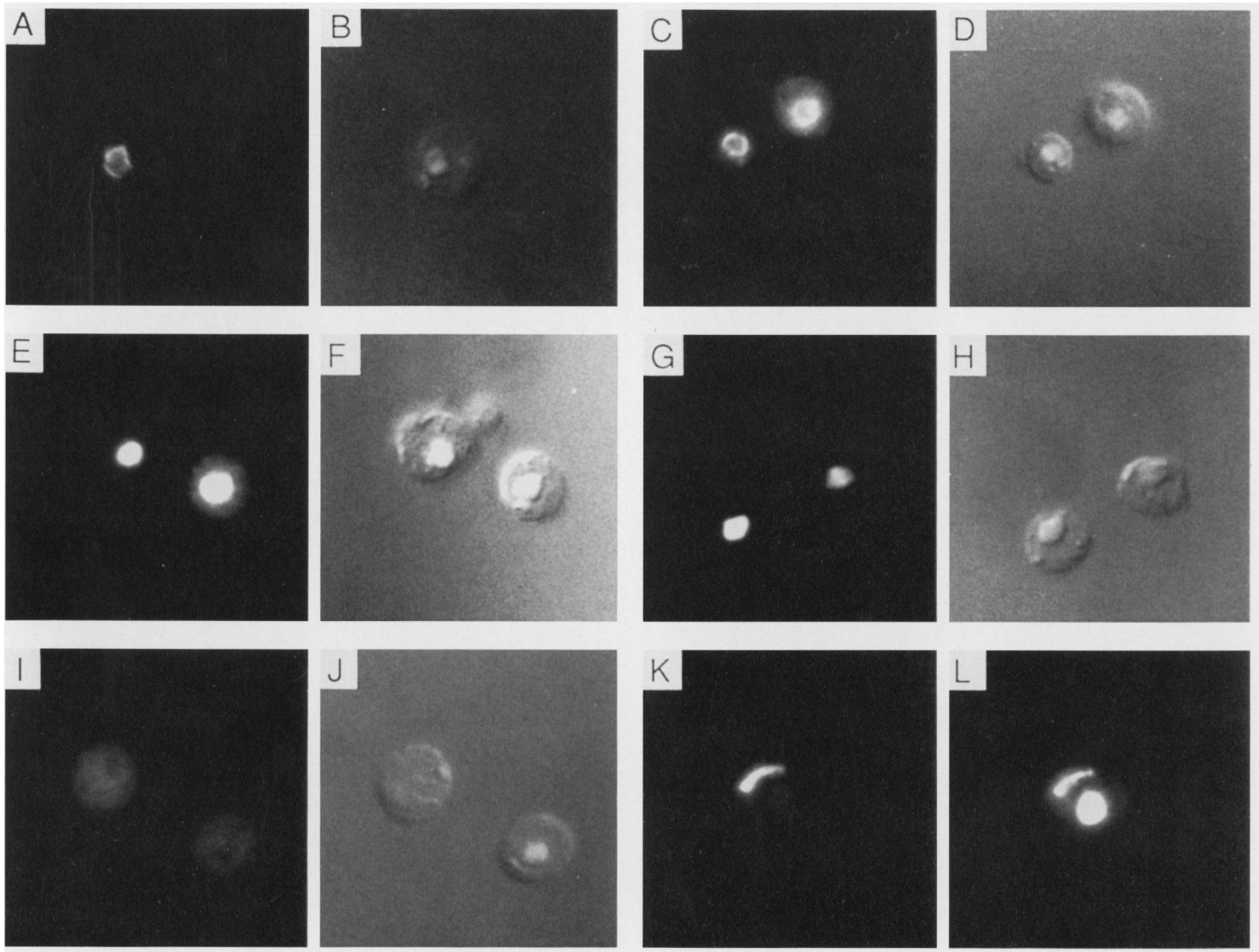


FIG. 8. Immunofluorescence localization of representative peptide fusion proteins. Yeast strain DBY2449 was transformed with derivatives of either pRB722 or pRB723. Cells were fixed and stained with either anti-invertase or anti-β-galactosidase followed by fluorescein isothiocyanate-conjugated secondary antibody and DAPI. 4H11 invertase was stained with anti-invertase antibody (A) and DAPI and Nomarski optics (B). 4H11–β-galactosidase was stained with anti-β-galactosidase antibody (C) and DAPI and Nomarski optics (D). 5G2-invertase was stained with anti-invertase antibody (E) and DAPI and Nomarski optics (F). 5G2–β-galactosidase was stained with anti-β-galactosidase antibody (G) and DAPI and Nomarski optics (H). 1D10 invertase was stained with anti-invertase antibody (I) and DAPI and Nomarski optics (J). 1C9 invertase was stained with anti-invertase antibody (K) and DAPI alone (L).

perinuclear region by immunofluorescence. The glycosylation state of these proteins indicates that they are persistently associated with the ER, suggesting that perinuclear staining is indicative of proteins located in the ER.

Several lines of evidence indicate that translocated proteins confined to the ER in yeast cells would reside in the nuclear envelope. Electron micrographs of yeast cells show the ER as membrane-bounded cisterna in contact with the plasma membrane and occasionally continuous with the two membranes of the nuclear envelope (17). As observed by freeze substitution electron microscopy, the ER and outer leaflet of the nuclear envelope are studded with ribosomes, as is the rough ER of mammalian cells (2). The nuclear envelope and connected ER membranes appear to be part of the same functional structure, and both have been shown to contain intracellular precursors of secreted acid phosphatase (31). The *KAR2* protein of yeast is both homologous to and functionally interchangeable with the mammalian heavy-chain-binding protein, an ER protein (22, 26). On the basis of immunofluorescence staining of yeast cells, the *KAR2* gene product is located at the nuclear periphery, further indicating that the nuclear envelope is part of the functional ER. Therefore, it is not surprising to find invertase and β-galactosidase hybrids that have glycosylation of ER proteins in association with the nuclear envelope. However, the invertase and β-galactosidase hybrids or *KAR2* protein are rarely seen at the cell periphery, and it is puzzling that these proteins appear to be located principally in the nuclear envelope and not in peripheral ER elements. Perhaps most of the ER membrane engaged in protein translocation in a yeast cell is in fact the outer membrane of the nuclear envelope. Alternatively, an appreciable fraction of the hybrid protein may be in an extranuclear network, but this material may be dispersed in a way that makes staining of this compartment less distinct than that of the nuclear envelope.

Immunofluorescence localization showed a diversity of possible fates of proteins fused to each of the 28 leader peptides examined here. To summarize, 10 of 28 sequences can direct invertase or β-galactosidase or both to a cellular compartment other than the cytoplasm. On the basis of the aggregate properties of hybrid proteins (Table 1), the peptides can be categorized as follows. Four of the peptides (4H11, 6C6, 6G8, and 8B5) are strong secretion signal sequences; they lead to glycosylation of either invertase or β-galactosidase and localize β-galactosidase to the nuclear perimeter. Three of the peptides (1B9, 1G10, and 2C3) are weak secretion signal sequences; they lead to low levels of glycosylated invertase and give rise to a significant fraction of invertase activity at the cell surface, but they do not lead to glycosylation of β-galactosidase, nor do they cause invertase or β-galactosidase to become localized to the nuclear perimeter. One of the peptides (5G2) is a strong nuclear localization signal; it causes either invertase or β-galactosidase to accumulate in the body of the nucleus but does not lead to glycosylation of either of these proteins. Four of the peptides (8B5, 2C3, 1C9, and 9G7) are mitochondrial localization sequences for invertase but not β-galactosidase; they lead to mitochondrial localization by immunofluorescence for invertase only. Two of these peptides (8B5 and 2C3) appear to be bifunctional localization signals, since they give rise to glycosylation of invertase as well as localization of invertase to mitochondria. The dual location of these hybrids could result from saturation of the translocation capacity of either the ER or mitochondria, diverting excess proteins to a secondary membrane target. However, it should be noted that invertase expression from the *PGK1* promoter is similar to induced expression from the wild-type *SUC2* promoter and is unlikely to exceed the translocation capacity of either the ER or mitochondria.

For the four leader peptides that direct invertase to the mitochondria, the corresponding peptide fusions to β-galactosidase do not show mitochondrial localization of β-galactosidase. This difference in behavior between the invertase and β-galactosidase fusion proteins suggests that invertase sequences contribute to the formation of a mitochondrial targeting sequence. β-Galactosidase itself has the capacity to be segregated to the mitochondria, since hybrid proteins composed of the mitochondrial signal sequence from the β subunit of the $F_1$ ATPase fused to β-galactosidase fractionate with mitochondria (10, 12).

A few percent of random fragments of *E. coli* DNA sequences can function as mitochondrial targeting signals for a nonfunctional cytochrome *c* oxidase subunit IV precursor lacking a presequence (3). Therefore, it is not surprising that mitochondrial targeting sequences were found in the collection of 28 random sequences examined here.

Several peptide sequences that direct proteins to the nucleus have been analyzed. These sequences are not similar to one another; however, they tend to be enriched for positively charged residues (9). Conceivably, the diversity of nuclear localization sequences reflects a low sequence specificity of recognition of these sequences as well. Our identification of 1 of 28 sequences as a nuclear localization signal is consistent with this hypothesis. It is possible that this sequence is an actual nuclear targeting signal derived from the coding sequence of a human nuclear protein, although the a priori probability of isolating such a sequence from a random library is very low.

In summary, these results imply that the segregation of proteins from the cytoplasm to different organelles is governed by peptide sequence recognition mechanisms that have very low sequence specificity. Targeting sequences that occur naturally have probably evolved to be recognized efficiently by only one of the sequence recognition mechanisms. It is only when these systems are presented with the wide diversity inherent in essentially random peptide sequences that the low specificity of recognition is revealed.

## ACKNOWLEDGMENTS

We thank Zhimin Zhu for DNA sequencing.

This work was supported by Public Health Service grants GM21253 and GM18973 from the National Institutes of Health and by a predoctoral fellowship from Robert Swanson to C.K.

## LITERATURE CITED

1. **Allison, D. S., and E. T. Young.** 1988. Single-amino-acid substitutions within the signal sequence of yeast prepro-α-factor affect membrane translocation. Mol. Cell. Biol. **8**:1915–1922.
2. **Baba, M., and M. Osumi.** 1987. Transmission and scanning electron microscope examination of intracellular organelles in freeze-substituted *Kloeckera* and *Saccharomyces cerevisiae* yeast cells. J. Electron Microsc. Technique **5**:249–261.
3. **Baker, A., and G. Schatz.** 1987. Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. Proc. Natl. Acad. Sci. USA **84**:3117–3121.
4. **Bankaitis, V. A., J. P. Ryan, B. A. Rassmussen, and P. J. Bassford.** 1985. The use of genetic techniques to analyze protein export in *Escherichia coli*, p. 105–150. *In* P. A. Knauf and J. S. Cook (ed.), Current topics in membranes and transport. Academic Press, Inc., New York.
5. **Bassford, P. J., Jr., T. J. Silhavy, and J. R. Beckwith.** 1979. Use

of gene fusion to study secretion of maltose-binding protein into *Escherichia coli* periplasm. J. Bacteriol. **139**:19–31.

6. **Bedouelle, H., P. J. Bassford, A. V. Fowler, I. Zabin, J. Beckwith, and M. Hofnung.** 1980. Mutations which alter the function of the signal sequence of the maltose-binding protein of *Escherichia coli*. Nature (London) **285**:78–81.
7. **Blachly-Dyson, E., and T. H. Stevens.** 1987. Yeast carboxypeptidase Y can be translocated and glycosylated without its amino-terminal signal sequence. J. Cell Biol. **104**:1183–1191.
8. **Blobel, G., and B. Dobberstein.** 1975. Transfer of proteins across membranes. I. Presence of proteolytically processed and nonprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. J. Cell Biol. **67**:835–851.
9. **Dingwall, C., and R. A. Laskey.** 1987. Protein import into the cell nucleus. Annu. Rev. Cell Biol. **2**:367–390.
10. **Douglas, M. G., B. L. Geller, and S. D. Emr.** 1984. Intercellular targeting and import of an $F_1$-ATPase b-subunit–β-galactosidase hybrid protein into yeast mitochondria. Proc. Natl. Acad. Sci. USA **81**:3983–3987.
11. **Emr, S. D., J. Hedgpeth, J. M. Clement, T. J. Silhavy, and M. Hofnung.** 1980. Sequence analysis of mutations that prevent export of lambda receptor, an *Escherichia coli* outer membrane protein. Nature (London) **285**:82–85.
12. **Emr, S. D., A. Vassarotti, J. Garrett, B. L. Geller, M. Takeda, and M. G. Douglas.** 1986. The amino terminus of the yeast $F_1$-ATPase beta-subunit precursor functions as a mitochondrial import signal. J. Cell Biol. **102**:523–533.
13. **Esmon, B., P. Novick, and R. Schekman.** 1981. Compartmentalized assembly of oligosaccharides on exported glycoproteins in yeast. Cell **25**:451–460.
14. **Hitzeman, R. A., F. E. Hagie, J. S. Hayflick, C. Y. Chen, P. H. Seeburg, and R. Deryuck.** 1982. The primary structure of the *Saccharomyces cerevisiae* gene for 3-phosphoglycerate kinase. Nucleic Acids Res. **10**:7791–7808.
15. **Kaiser, C. A., and D. Botstein.** 1986. Secretion-defective mutations in the signal sequence for *Saccharomyces cerevisiae* invertase. Mol. Cell. Biol. **6**:2382–2391.
16. **Kaiser, C. A., D. Preuss, P. Grisafi, and D. Botstein.** 1987. Many random sequences functionally replace the secretion signal sequence of yeast invertase. Science **235**:312–317.
17. **Matile, P., H. Moor, and C. Robinow.** 1969. Yeast cytology, p. 219–302. *In* A. H. Rose and J. S. Harrison (ed.), The yeasts, vol. 1. Academic Press, Inc., New York.
18. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
19. **Michaelis, S., J. F. Hunt, and J. Beckwith.** 1986. Effects of signal sequence mutation on the kinetics of alkaline phosphatase export to the periplasm in *Escherichia coli*. J. Bacteriol. **167**: 160–167.
20. **Milstein, C., G. G. Brownlee, T. M. Harrison, and M. D. Mathews.** 1972. A possible precursor of immunoglobulin light chains. Nature (London) New Biol. **239**:117–120.
21. **Ngsee, J. K., W. Hansen, P. Walter, and M. Smith.** 1989. Cassette mutagenic analysis of the yeast invertase signal peptide: effects on protein translocation. Mol. Cell. Biol. **9**:3400–3410.
22. **Normington, K., K. Kohno, Y. Kozutsumi, M.-J. Gething, and J. Sambrook.** 1989. S. cerevisiae encodes an essential protein homologous in sequence and function to mammalian BiP. Cell **57**:1223–1236.
23. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
24. **Preuss, D., and D. Botstein.** 1989. Intragenic revertants of yeast variants with secretion-defective leader sequences. Mol. Cell. Biol. **9**:1452–1464.
25. **Randall, L. L., S. J. S. Hardy, and L. G. Josefsson.** 1978. Precursors of three exported proteins of *E. coli*. Proc. Natl. Acad. Sci. USA **75**:1209–1212.
26. **Rose, M. D., L. M. Misra, and J. P. Vogel.** 1989. KAR2, a karyogamy gene, is the yeast homolog of mammalian BiP/GRP778. Cell **57**:1211–1221.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
28. **Schauer, I., S. Emr, C. Gross, and R. Scheckman.** 1985. Invertase signal and mature sequence substitutions that delay intercompartmental transport of active enzyme. J. Cell Biol. **100**: 1664–1675.
29. **Stadner, J., S. A. Benson, and T. J. Silhavy.** 1986. Kinetic analysis of *lamB* mutants suggests the signal sequence plays multiple roles in protein export. J. Biol. Chem. **261**:15075–15080.
30. **Trimble, R. B., and F. Maley.** 1977. Subunit structure of external invertase from Saccharomyces cerevisiae. J. Biol. Chem. **252**:4409–4412.
31. **van Rijn, H. J. M., W. A. M. Linnemans, and P. Boer.** 1975. Localization of acid phosphatase in protoplasts from *Saccharomyces cerevisiae*. J. Bacteriol. **123**:1144–1149.
32. **Walter, P., and V. R. Lingappa.** 1986. Mechanism of protein translocation across the endoplasmic reticulum membrane. Annu. Rev. Cell Biol. **2**:499–516.