Many Random Sequences Functionally Replace the Secretion Signal Sequence of Yeast Invertase

CHRIS A. KAISER, DAPHNE PREUSS, PAULA GRISAFI, DAVID BOTSTEIN

In the process of protein secretion, amino-terminal signal sequences are key recognition elements; however, the relation between the primary sequence of an amino-terminal peptide and its ability to function as an export signal remains obscure. The limits of variation permitted for functional signal sequences were determined by replacement of the normal signal sequence of *Saccharomyces cerevisiae* invertase with essentially random peptide sequences. Since about one-fifth of these sequences can function as an export signal the specificity with which signal sequences are recognized must be very low.

THE STUDY OF THE MOVEMENT OF PROTEINS FROM THE cytoplasm through membranes has been guided by the fundamental principle that each protein carries within its amino acid sequence the information specifying its proper cellular location. The most prominent class of peptide sequences that govern protein location is the amino-terminal "leader" or "signal" sequences found in precursors of exported proteins of bacteria and higher cells. One way to define the mechanism by which these sequences are recognized is to investigate the specific elements within the signal peptide that are required for function. Progress in this area has come from sequence comparisons and mutational analysis. For example, one of the features that the more than 200 known signal sequences have in common is a stretch of at least seven hydrophobic amino acid residues (*1*). This hydrophobic region is believed to be critical for signal function since mutations in bacterial secretory proteins that reduce the length of the hydrophobic region either by deletion or by the insertion of charged residues generally lead to a severe block in export (*2*). The role of different signal sequence elements in the export of eukaryotic proteins is less clear. Although mutations have been isolated in the signal sequences of eukaryotic proteins that disrupt export function (*3–6*), all are either deletions or substantial rearrangements.

To analyze the fine structure of the signal sequence of *Saccharomyces cerevisiae* invertase, we attempted to isolate secretion-defective point mutations. Even when the signal sequence region was subjected to localized in vitro mutagenesis we failed to isolate mutants defective in invertase export (*7*). In a related experiment Ngsee and Smith (*8*) generated and recovered many point mutations by replacement of the invertase signal sequence with synthetic DNA containing random nucleotide substitutions. A large fraction of these mutations produce secreted and fully glycosylated invertase even though they demonstrably introduce one or more charged residues into the hydrophobic region of the signal. These findings show that many variant leader sequences that differ from wild type as a result of a small number of amino acid changes retain at least partial signal function. This raises the pivotal question of just how many alternative sequences can function as a signal sequence in yeast.

We present here a direct measurement of the proportion of polypeptides of essentially random amino acid sequences that can act as an export signal in yeast. We find that an unexpectedly large fraction (20 percent) of such sequences will direct invertase export, showing that the specificity involved in signal sequence recognition is very low indeed.

Isolation of functional signal sequences. The *SUC2* gene encodes two forms of invertase: a constitutive cytoplasmic enzyme, and a secreted glycosylated enzyme that is regulated by glucose repression. The protein moieties of the two forms of invertase differ only in that the secreted form is synthesized with an additional amino-terminal signal sequence of 20 hydrophobic amino acids that is cleaved during invertase export (*9–11*). The plasmid pRB576 was designed to allow the replacement of the wild-type *SUC2* signal sequence with sequences encoded by inserted DNA fragments. Replacement of the signal sequence is made possible by the *suc2-450* allele which is a deletion that extends from the third to the twentieth codon (the normal signal processing site is after amino acid nineteen) and an inserted Sma I linker that encodes three additional amino acids (Phe-Pro-Gly) not found in the wild-type sequence (Fig. 1). This deletion allele itself gives rise to the production of an active, intracellular, nonglycosylated invertase (Figs. 2 and 5 and Table 1). A similar deletion allele has been previously characterized in detail, and the enzyme produced was shown to be soluble and cytoplasmic (*5*). We have inserted a large number of different DNA fragments of essentially random sequence at the position of the Sma I site in order to identify those that can substitute for the wild-type sequence as an export signal. Human genomic DNA digested with the restriction enzymes Hae III, Rsa I, and Hinc II was used as a source of DNA fragments with high complexity. A library of fragments ligated into the Sma I site was generated, and the DNA was amplified in pools of *Escherichia coli* transformants and was introduced by transformation into yeast.

The secreted form of invertase is required for the utilization of sucrose by yeast such that strains that are not capable of producing the secreted enzyme grow poorly on media containing sucrose as the sole carbon and energy source. The expression and regulation of a wild-type *SUC2* allele carried on pRB576 are equivalent to that of a chromosomal copy of the gene and allow growth on sucrose (*5*), whereas the *suc2-450* allele under the same conditions allows only very weak growth on sucrose (Table 1). Therefore, transformants that produce invertase with a functional signal sequence can be identified simply by their ability to grow on sucrose plates. When a yeast strain deleted for the chromosomal copy of the *SUC2* gene was transformed with DNA from the library and plated on medium

The authors are in the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

containing sucrose as the carbon source, 0.8 percent (164/19,610) of the transformants grew well. If the invertase molecules produced by these colonies have reached the cell surface via the secretory pathway they should be glycosylated. Extracts prepared from 40 of the Suc^+ transformants were resolved by native polyacrylamide gel electrophoresis (PAGE), which readily separate glycosylated from nonglycosylated invertase. All of these strains produce active glycosylated enzyme (Fig. 2A), as anticipated.

The frequency of inserts that lead to the production of invertase, regardless of their ability to function as export signals, can be estimated directly. If we assume that the sequences of bulk human DNA are essentially random, on average only one-third of the inserted sequences will be a multiple of three base pairs in length and therefore maintain the correct reading frame of the *SUC2* gene. Furthermore, an appreciable proportion of the inserted sequences will contain termination codons. Based on the size of 22 inserts chosen at random (median length, 200 base pairs) the frequency of inserts expected to be in frame with no termination codons is about 4 percent (*12*). Therefore, it appears that, within the class of invertase amino-terminal peptides encoded by this library, a remarkably large fraction (0.8/4.2, about 20 percent) function, at least partially, as export signals.

As an independent means of estimating the frequency of inserts that function as signal sequences, a collection of transformants were screened for the ability to produce any form of invertase by gel immunoblotting with antiserum to invertase. This method will identify transformants that produce invertase regardless of their ability to grow on sucrose. We found that a large fraction (about one-third) of the transformants constitutively produce a nonglycosylated form of invertase with the mobility of wild-type cytoplasmic invertase. These are cases where the inserted DNA has restored promoter activity (deleted in *suc2-450*) thereby restoring expression of the normal constitutive internal form. To avoid confusing this form of the enzyme with cytoplasmic forms of invertase containing an altered leader peptide, cell extracts were resolved by SDS-PAGE, which allows the normal cytoplasmic form of invertase and forms with as few as nine additional amino acids to be distinguished (*5*). Out of 407 clones screened by gel immunoblotting, 14 produce forms of invertase that are not glycosylated and have perceptibly lower mobilities than the normal cytoplasmic enzyme (for example see Fig. 5, *suc2-935*). As expected, all of these isolates fail to produce detectable glycosylated forms of invertase when assayed on a native gel (Fig. 2B). Three of the 407 transformants were able to grow on sucrose which is consistent with the value of 0.8 percent found by screening a much larger number of clones. Of these three, two clearly produced glycosylated invertase. The third did not produce any invertase detectable against the background in the immunoblot, and we assume that this strain makes only a small amount of the secreted enzyme. Thus the frequency of transformants in this sample that produce a form of invertase with an additional amino-terminal peptide is 4.2 percent, and again about 20 percent (3/17) of these amino-terminal peptides function as signals. Overall these results fully support the frequency estimates based on the distribution of insert sizes.

The transformed yeast that grow on sucrose plates form a continuous distribution of colony sizes; as the plates are incubated, colonies continue to arise until the eventual yield of slow growing colonies is comparable to that of the colonies that are clearly Suc^+. Approximately half of the slow growing transformants produce high levels of nonglycosylated invertase, and most of these appear to be the *suc2-450* allele without an insert. The others produce low levels of glycosylated invertase, which is detectable on an activity gel. Since this second group was not counted in our estimation of the frequency of Suc^+ transformants, 0.8 percent is likely to be an underestimate of the true frequency of clones able to produce some glycosylated invertase. Therefore 20 percent is an underestimate of the fraction of sequences that can function at least minimally as signal sequences.

Structure of leader sequences. The DNA sequence of representatives of both the functional and nonfunctional class of leader sequences was determined and the derived amino acid sequences are shown in Fig. 3. All of the sequences from independently isolated plasmids are different. The altered forms of invertase expressed by all of these strains were found to be regulated by glucose, an indication that translation initiates at the same methionine codon as for the

Table 1. Functionality of random amino-terminal sequences in secretion of invertase.

SUC2 allele	Total* activity (units)	Extracellular† activity (%)	Glycosylation‡		Leader processing§	Growth‖ on sucrose
			Amount (%)	Type		
SUC2	1260	94	>80	OC	Y	+++++
450	1060	1.9	<20		N	+
suc2Δ	7	0.0	–			–
317	260	1.6	<20			–
527	56	1.4	<20			–
546	1890	0.6	<20			+
615	750	1.3	<20			+
645	1130	1.4	<20			+
740	1300	0.9	<20			+
756	64	1.1	<20			–
827	23	0.7	<20			–
935	1270	0.9	<20		N	+
102	1060	9.0	>80	C, OC	N	++++
201	760	89	>80	OC	Y	+++++
203	620	57	>80	C, OC	Y/N	+++++
204	400	34	>80	C	N	++++
301	512	82	>80	OC	N	+++++
310	310	2.6	20–80	C	N	++
401	370	52	20–80	OC	N	+++++
402	790	11	20–80	OC		++++
501	550	68	>80	C	N	+++++
502	750	63	>80	C, OC	N	+++++
503	1490	23	>80	C	N	+++++
601	670	33	>80	C, OC	N	+++++
701	420	66	20–80	C, OC		+++++
704	190	62	>80	OC	N	++++
801	110	51	<20			+++++
802	930	52	>80	C	N	+++++
809	340	3.4	<20			+++
810	270	20	<20			++++
901	1080	39	>80	C, OC,×	N	+++++
903	370	38	>80	C		+++++
909	300	4.3	20–80	C		++
1001	360	12	<20			++++
1010	760	1.0	20–80	C		+

*Cultures of strain YT455 carrying various *SUC2* alleles were grown in YEP medium with 2 percent lactate (*p*H 5.5) and 0.1 percent glucose to exponential phase. The total activity is the sum of the activity in the culture supernatant, on the surface of intact cells, and in the spheroplast lysate. Spheroplasts and intact cells were prepared as described (*5*), and invertase was assayed by the method of Goldstein and Lampen (*20*). Invertase units are nanomoles of glucose released per minute per absorbancy unit (A_{600nm}) at 37°C. The values are averages of two determinations that did not differ by more than 20 percent. The allele designated *suc2Δ* is YT455 without a plasmid. †The fraction of the total activity that is in the culture supernatant and on the cell surface. For all strains, less than half of the external activity is in the medium. ‡The intensities of the glycosylated and nonglycosylated forms of invertase produced after a 3-hour induction period were compared after SDS-PAGE and immunoblotting. The types of glycosylation are defined as follows: C, core—discrete bands approximately 90 kD; OC, outer chain—diffuse bands more than 110 kD; and ×, a novel form that is probably an invertase dimer. §Leader processing was assessed by the comparison of the size of the protein moiety of mutant invertase after treatment with endo H or induction in the presence of tunicamycin with that of the protein moiety of processed wild-type invertase (Y, yes; N, no). Strains that produce low amounts of glycosylated invertase were not tested. ‖About 10^6 cells were spotted on solid YEP medium containing 2 percent sucrose and Antimycin A at 1 μg/ml. Relative growth was scored after incubation for 2 to 3 days at 30°C.

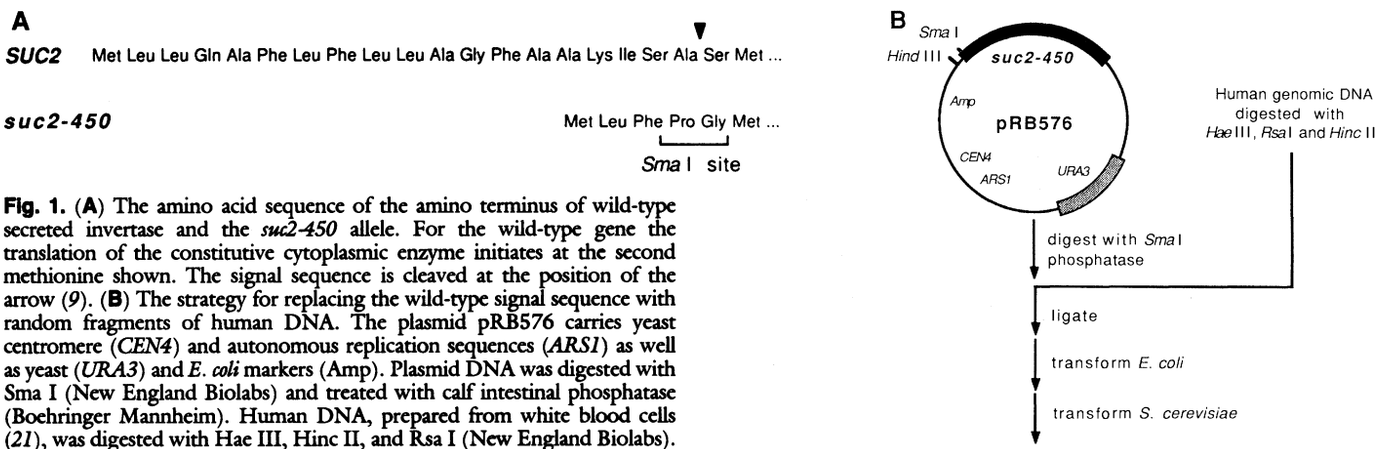

Fig. 1. (**A**) The amino acid sequence of the amino terminus of wild-type secreted invertase and the *suc2-450* allele. For the wild-type gene the translation of the constitutive cytoplasmic enzyme initiates at the second methionine shown. The signal sequence is cleaved at the position of the arrow (*9*). (**B**) The strategy for replacing the wild-type signal sequence with random fragments of human DNA. The plasmid pRB576 carries yeast centromere (*CEN4*) and autonomous replication sequences (*ARS1*) as well as yeast (*URA3*) and *E. coli* markers (Amp). Plasmid DNA was digested with Sma I (New England Biolabs) and treated with calf intestinal phosphatase (Boehringer Mannheim). Human DNA, prepared from white blood cells (*21*), was digested with Hae III, Hinc II, and Rsa I (New England Biolabs). The ligation mixture contained 5 μg of vector DNA, 0.5 μg of digested human DNA and T4 DNA ligase (New England Biolabs). After incubation at 25°C, the DNA was diluted into buffer and digested with Sma I to linearize vector DNA with no insert. The *E. coli* strain DB6507 was transformed with the ligated DNA, and ampicillin-resistant clones were selected. Ten pools of about 1000 transformants each were collected. Restriction enzyme mapping of plasmids from random transformants revealed that 21 of 22 had an inserted fragment. Yeast strain YT455 (a *suc2-Δ9, ura3-52,* and *ade2-101*) was transformed with plasmid DNA from each pool (*22*). Ura$^+$ transformants were selected by growth on solid medium containing yeast nitrogen base (Difco), 2 percent glucose, adenine at 50 μg/ml and 2 percent agar. The transformants were either transferred to solid YEP medium (1 percent yeast extract, 2 percent peptone, and 2 percent agar) with 2 percent sucrose and Antimycin A (Sigma) at 1 μg/ml to screen for the ability to grow on sucrose or were screened for the production of invertase by immunoblotting (*23*).

wild-type secreted enzyme. For all of the sequences with the exception of *suc2-827* (*13*), this methionine codon is in frame with the *SUC2* coding sequence; and in all but six cases this is the only methionine codon in the correct reading frame.

A comparison of the functional and nonfunctional sequences should tell us something of the rules that govern signal sequence function. (For this comparative analysis, the presence of glycosylated invertase on an activity gel or immunoblot, rather than growth on sucrose, was used as the ultimate criterion for function as an export signal.) An obvious distinction between the two classes is the tendency for functional sequences to be hydrophobic. As a group the hydrophobic residues (Phe, Ile, Leu, Val, and Met) are enriched 2.6-fold and charged residues (Glu, Asp, Lys, and Arg) are depleted 2.8-fold in the functional class. This suggests that signal function may be based on a property of the leader sequence that is related to hydrophobicity. The ability of an individual sequence to function as a signal is clearly correlated with the hydrophobicity and density of charged residues of that sequence (Fig. 4A). The same criteria can be applied to natural yeast signal sequences and the amino termini of yeast cytoplasmic proteins (Fig. 4B). It is clear that these naturally occurring functional and nonfunctional amino-terminal sequences are separated and that the boundary between the two classes falls roughly at the same position as for the sequences analyzed in this study. The simplest explanation of these data is that generally any amino-terminal peptide that has a hydrophobicity above some threshold value and has few charged residues will function at least partially as a signal sequence.

Characterization of exported invertase. Wild-type invertase transits the secretory pathway, undergoing both core glycosylation and addition of outer chains to yield a heterodisperse population of extracellular glycoproteins with an average molecular mass of about 120 kD (*14*). Many of the hybrid invertase molecules characterized in this study differ from the wild-type enzyme in both cellular location and structure. To examine a broad range of phenotypes, individual strains representing both the fast and slow growing categories as well as some with inserts that clearly do not function as signal sequences were characterized in detail.

The extracellular (medium and cell surface) and intracellular

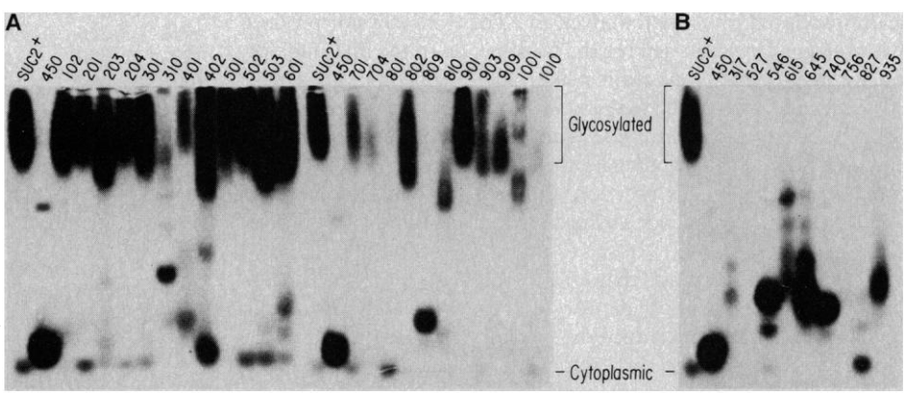

Fig. 2. Native polyacrylamide gel showing the different forms of invertase specified by representative alleles. (**A**) Alleles isolated by the ability to allow growth on sucrose plates; and (**B**) alleles that produce nonglycosylated invertase isolated in the immunoblotting screen. Alleles whose numbers end with 10 or 09 (for example *suc2-310* and *suc2-909*) were isolated as colonies that grew slowly on sucrose. The position of the normal cytoplasmic and the secreted forms of invertase are indicated. The different phenotypes of the strains isolated in the two screening methods were shown to be due to the plasmidborne *suc2* allele by reisolation of each plasmid; plasmid DNA was recovered from each yeast strain (*24*) used to transform *E. coli*, and then DNA prepared from *E. coli* was reintroduced into the *suc2Δ* yeast strain YT455 by transformation (*22*). Cultures were grown to the exponential phase in liquid YEP with 2 percent lactate (*p*H 5.5) and 0.1 percent glucose to allow steady state synthesis of invertase. Cells (2×10^7) were lysed by agitation in the presence of 20 μl of buffer [100 m*M* tris-phosphate, *p*H 6.7, 10 percent glycerol, 0.05 percent bromphenol blue, 1 m*M* phenylmethylsulfonyl fluoride (PMSF)] and 0.1 g of glass beads (0.5 mm). More buffer was added (30 μl), the cell debris was removed by centrifugation, and 20 μl of this crude extract was placed on each lane of a 5 percent polyacrylamide gel with 100 m*M* tris-phosphate (*p*H 6.7) buffer. Electrophoresis was conducted at 4°C overnight, and the gels were stained for invertase activity (*25*). The glycosylated forms of *suc2-801* and *suc2-809* can be seen if more extract is loaded on the gel.

(spheroplast lysate) enzyme activities were measured independently (Table 1). The results are consistent with the expectation that only strains that express glycosylated invertase will be able to produce appreciable levels of extracellular enzyme and to grow well on sucrose. All of the strains isolated by immunoblotting that do not produce glycosylated invertase have very low levels of extracellular invertase activity. In contrast, the strains that produce glycosylated invertase generally have a significantly higher fraction of extracellular activity, which in some cases (*suc2-201* and *suc2-301*) approaches that of the wild-type enzyme. For most of these strains, however, a significant fraction of the enzyme is intracellular even in cases where essentially all of the invertase is glycosylated. This suggests that the altered signal sequences give rise to the intracellular accumulation of glycosylated enzyme.

Previous studies have shown that mutations in secretory proteins that block signal sequence cleavage lead to a dramatic reduction in the rate of transport of these proteins to the cell surface. Such mutations have been isolated in the *SUC2* (*15*) and *PHO5* (*16*) genes. One of the mutations in *SUC2* lies at the leader peptidase cleavage site, and hence it blocks the cleavage of the signal peptide and causes invertase to accumulate in the endoplasmic reticulum (ER) in a core glycosylated state. To test whether failure of signal cleavage is the basis of the accumulation of intracellular invertase in our study, we used immunoblotting to investigate the glycosylation state and signal processing of forms of invertase produced by each of the strains. The predominant form of invertase produced by most of the Suc^+ transformants has a molecular mass of approximately 90 kD (*suc2-503,* for example), an indication that these enzymes have received core glycosylation only and have not reached the Golgi, the site of outer chain addition (*14*). The size of the protein moiety alone was examined by either specific enzymatic removal of carbohydrate with endo-β-*N*-acetylglucosaminidase H (endo H) or by blocking carbohydrate addition with tunicamycin. Both treatments gave rise to a form of invertase with a significantly lower mobility than the corresponding form of wild-type invertase, showing that additional peptide sequences are present. Therefore, a simple failure of signal cleavage can explain the intracellular accumulation of invertase exhibited by these *SUC2* alleles.

One exceptional strain exhibits apparently normal processing of the signal peptide. The protein moiety of the glycosylated invertase protein specified by *suc2-201* is the size of the wild-type processed enzyme. Furthermore, this glycosylated enzyme appears to have outer chains added and to reach the cell surface efficiently, supporting the view that removal of the signal peptide facilitates transport to the cell surface. The protein moiety of invertase specified by *suc2-203* has a lower apparent molecular size than expected for a protein that has retained the full-length signal (29 amino acids), indicating that this protein is also processed. However, cleavage in this case does not occur close to the normal processing site since the protein has a greater molecular mass than does processed wild-type protein. This emphasizes that we cannot yet tell whether processing is by the normal route for any altered protein.

Some of the altered signal sequences appear to influence invertase maturation in unexpected ways. First, the invertase produced by *suc2-401, suc2-402*, and *suc2-704* appears to have received outer chains; nevertheless, a large proportion of the enzyme remains intracellular. This indicates an unusual kinetic block in secretion that occurs after outer chain addition. Second, the *suc2-901* allele specifies the production of an unusual form of invertase which has a much greater apparent molecular size than that of the fully glycosylated wild-type enzyme (Fig. 5). Preliminary experiments indicate that this form is a dimer of invertase subunits linked by an interaction between the leader peptides that is not dissociated by boiling in the presence of SDS and a sulfhydryl reducing agent.

Immunoblotting was also used to estimate the efficiency membrane translocation of invertase expressed by different alle The relative amounts of glycosylated and nonglycosylated inver produced during a 3-hour derepression gives an estimate of fraction of enzyme that has reached the lumen of the ER within

Fig. 3. The derived amino acid sequences of functional and nonfunctic leader peptides. Each sequence begins with the initiator methionine of normally secreted form of invertase [except for *suc2-827* (*13*)] and ends v the initiator methionine of cytoplasmic invertase. The hydrophobic resic (Met, Phe, Ile, Leu, Val) are underlined and the charged residues indicated. The DNA sequence was derived from plasmid DNA carry different leader sequences that was labeled at the 3′ end of a Hind III which is 26 bp upstream of the initiator codon for the secreted form of w type invertase. The sequencing method of Maxam and Gilbert (*26*) was u and an alkaline hydrolysis reaction (A > C) was carried out in addition to four standard base modification reactions. The sequence of only one str was determined. The single-letter abbreviations are as follows: A, Ala Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, M N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, T

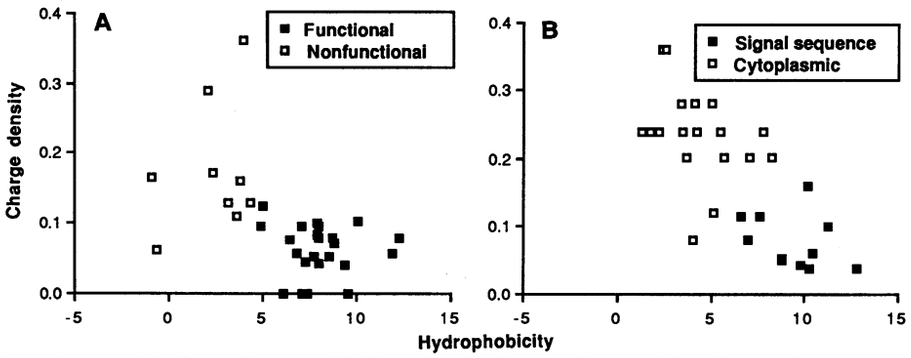

Fig. 4. A plot showing the hydrophobic properties of various amino-terminal sequences. For each sequence, the total hydrophobicity of the most hydrophobic ten-amino-acid segment is plotted against the density (per residue) of charged amino acids (Arg, Lys, Asp, Glu). The hydrophobicity scale of Kyte and Doolittle (*27*), normalized to a mean of 0.0 and a standard deviation of 1.0, was used. Similar results are obtained with the hydrophobicity scale of Engleman, Steitz, and Goldman (*28*) or with a window size of 5 or 15 (rather than 10). (**A**) Comparison of the functional and nonfunctional sequences isolated. The actual sequences analyzed are those shown in Fig. 3. (**B**) Comparison of the signal sequences of known yeast secretory proteins with the amino termini of a collection of yeast cytoplasmic proteins (that is, not secreted). The actual sequences analyzed are the putative signal sequences of the yeast secretory and membrane proteins encoded by the following genes: *MFα1, MFα2, PHO5, SUC2, KEX2, PRC1, STE3, BAR1, KIL-K, STA1, MEL1*, and the first 25 residues of the following yeast cytoplasmic proteins: alcohol dehydrogenase II, glyceraldehyde 3-phosphate dehydrogenase, superoxide dismutase (CuZn), phosphoglycerate kinase, anthranilate synthase component I, anthranilate synthase component II, ribosomal protein S33, ribosomal protein 51, adenosine triphosphate phosphoribosyltransferase, pyruvate kinase, thymidylate kinase, phosphoribosyl–adenosine monophosphate cyclohydrolase, enolase, tryptophan synthase, triosephosphate isomerase, phosphoglyceromutase, methionyl–transfer RNA synthetase, elongation factor 1-αA (*29*).

induction period, provided that no significant degradation of invertase occurs. For each strain the approximate amount of induced enzyme that is glycosylated is shown in Table 1. Perhaps surprisingly, many of the mutant proteins appear to be translocated as efficiently as the wild-type enzyme, at least within the limits of resolution afforded by this experiment. Overall there is a broad range of efficiencies and the invertase produced by *suc2-310* is shown in Fig. 5 as an example of an allele that has both a low level of expression and a low efficiency of translocation.

Generality and implications of the low information content of signal sequences. We have shown above that the normal signal sequence of yeast invertase can be replaced by any of a number of unrelated sequences that will function as export signals. About 20 percent of sequences derived at random from the human genome will direct invertase export, as judged by both growth on sucrose as a carbon and energy source and glycosylation of the enzyme. The large number of sequences that satisfy the criteria for recognition as an export sequence means that the information content of the leader peptide is very low, so low as to suggest that some gross property shared to some degree by one-fifth of all random sequences is recognized as the export signal.

Before discussing the implications of this result, we address some possible concerns about its generality. First, we can argue that the outcome of this experiment would not be different if some source of random DNA sequences other than the human genome had been used. The fraction of human genomic DNA that actually serves as coding sequence is well below 20 percent (*17*), and the fraction that encodes authentic signal peptides is much lower still. Thus, it is exceedingly unlikely that any of the human sequences isolated in this experiment have evolved as signal sequences. Analysis of the sequences themselves supports the contention that they approach randomness. The frequency of hydrophobic residues (Phe, Leu, Ile, Val, and Met) encoded by the functional and nonfunctional inserts together is 20 percent (calculated as a weighted mean to account for the observed fourfold larger abundance of nonfunctional sequences), a value very close to the 26 percent expected for random DNA (that is, all codons equally likely). Likewise the frequency of charged residues for the inserts and for the codon table is the same (20 percent).

The second area of concern is the premise that the region that we deleted and replaced is indeed the signal sequence for invertase. There are few sequences whose in vivo role as an export signal is more certain than the first 20 residues of invertase. This sequence satisfies all of the standard criteria for signal sequences: it is the only difference between the coding sequences of the internal and secreted forms (*9–11*), it is processed during export (*9*), and it is necessary and sufficient for association with the ER. Specifically, we (*5*) and others (*6*) have shown that deletions such as *suc2-450*, which remove most or all of the signal sequence, result in normally regulated synthesis of a totally cytoplasmic, nonglycosylated form of the active enzyme. These findings, as well as the original observation that the normal cytoplasmic enzyme results from the synthesis of a form of invertase without the signal sequence, show that the signal sequence

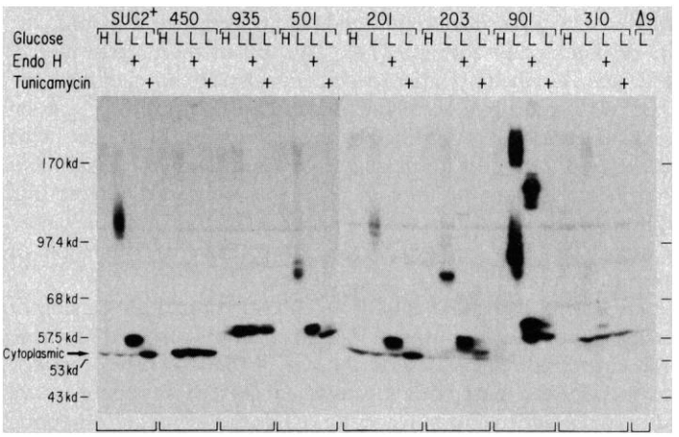

Fig. 5. Immunoblot showing the regulation, glycosylation state, and leader processing of invertase specified by various signal sequence alleles. The position of the constitutive cytoplasmic form of invertase is indicated. Yeast strain YT455 carrying a derivative of pRB576 was grown to exponential phase in YEP with 5 percent glucose. Cells (about 2×10^7) were centrifuged and suspended in either repressing medium, YEP with 5 percent glucose (H), or derepressing medium, YEP with 0.1 percent glucose (L). Tunicamycin was added at 10 μg/ml where indicated. The cultures were incubated at 30°C for 3 hours and washed once with 25 m*M* tris, *p*H 7.5, and 10 m*M* sodium azide. Cell pellets were lysed by agitation on a Vortex mixer in the presence of 20 μl of sample buffer (80 m*M* tris, *p*H 6.8, 2 percent SDS, 0.01 percent bromphenol blue, 0.1*M* dithiothreitol, 10 percent glycerol, 2 m*M* PMSF) and 0.1 g of glass beads, heated at 95°C for 2 minutes, and centrifuged at 12,000*g* to remove cell debris. For endo H digestion, the samples were diluted with three volumes of 20 m*M* sodium citrate, *p*H 5.5; endo H (25 U/mg) was added to a final concentration of 3 μg/ml, and the mixture was incubated at 37°C for 18 hours. The other samples were diluted with three volumes of sample buffer. A portion (15 μl) of each (except for extracts from *suc2-310* where 25 μl were used) was resolved by SDS-PAGE on a 7.5 percent gel (*30*), and invertase was identified by immunoblotting (*5*). The difference in mobility of the forms produced by endo H and tunicamycin treatment can be accounted for by the residual *N*-acetylglucosamine residues that remain after endo H treatment (*31*). The faint bands at 110 kD, 90 kD, and 67 kD are probably not related to invertase since they are present in YT455, which is deleted for the *SUC2* gene (last lane).

is necessary for secretion. Furthermore, the signal alone appears to contain sufficient information to direct a cytoplasmic protein to the ER in that Emr *et al.* demonstrated that β-galactosidase, when fused to the first 65 residues of invertase, becomes associated with the ER membrane (*18*). We have shown in a similar fusion experiment that the first 19 amino acids of invertase suffice to cause core glycosylation of β-galactosidase (*7*). Thus, most if not all of this critical element specifying invertase secretion has been deleted in the *suc2-450* allele.

However, we cannot and do not wish to exclude the possibility that there is additional information in the body of the invertase structural gene that aids in proper localization. All we can say is that such additional information is insufficient in itself since the deletion protein specified by *suc2-450* as well as nonfunctional insert-carrying proteins remain cytoplasmic. We conclude that the functional inserts do indeed specify sequences that act, albeit with varying efficiency, as substitutes for the normal secretion signal.

If we assume that the translocation of invertase into the ER is initiated by the binding of the signal sequence to some sort of receptor molecule, our results place considerable constraints on the specificity with which the interaction can occur. The interaction with a receptor would have to be of a specificity low enough to account for the fact that roughly one-fifth of all peptide sequences can interact productively to some degree. This result argues against extensive specific contacts with particular amino acid side chains as the basis of interaction. The tendency for functional signals to be hydrophobic suggests that the signal binding site may be an apolar groove on the surface of the receptor molecule that will accommodate virtually any hydrophobic peptide of 10 to 20 residues.

An interesting outcome of these findings is the demonstration that the stretch of seven or more hydrophobic amino acids, common to all naturally occurring signal sequences, is not absolutely required for signal function. For example, two of the sequences analyzed in detail, *suc2-310* and *suc2-501*, have no more than three contiguous hydrophobic amino acids that are not interrupted by either charged or large polar residues. Furthermore, a basic amino terminus, commonly found in signal sequences, is also not found in many of the functional sequences here. Yet these features of natural leader sequences have evolved, and therefore we can reasonably ask what role they might play in leader sequence function.

Some specific possibilities are suggested by the properties of the leader sequences analyzed here. First, native signal sequences might have evolved to be particularly efficient in the initiation of translocation. To the first approximation, this hypothesis is not supported by our data since alleles such as *suc2-501* appear to be translocated as efficiently as the wild type, on the basis of the methods used here. However, a more detailed kinetic analysis of translocation might reveal distinctions between the sequences isolated in our study and the wild-type signal. In considering the types of selective pressures contributing to the conservation of certain leader sequence elements, it is important to allow for the possibility that leader sequences have evolved to perform functions beyond simply being recognition elements for membrane translocation. It might therefore be significant that only two of the signal peptides examined in our study appear to be cleaved by leader peptidase yet most of them contain at least one sequence that fits the empirical rules for a peptidase cleavage site (*19*). Conceivably, determinants within the signal peptide, in addition to those at the actual cleavage site, govern recognition by leader peptidase. Finally, it is likely that leader peptides play a role in gene expression since various leader sequences encoded by human DNA give rise to different levels of invertase protein. This effect appears to be independent of protein export since both functional and nonfunctional classes of leader show variability in the level of invertase. Conceivably then, a long hydrophobic domain and basic residues at the amino terminus are maintained in natural signal sequence because these sequence elements play a role in efficient translocation, peptidase recognition, or efficient gene expression.

The power of our method lies in its ability to separate these different functions of the amino-terminal peptide. By only demanding the production of a low level of extracellular invertase, we have isolated alleles of *SUC2* that have different efficiencies of invertase expression, translocation into the ER, and transport to the cell surface. The ability of different peptide sequences to influence each of these processes can now be directly compared. As more sequences are isolated and analyzed, the relation between amino acid sequence and each of these functions will be clarified.

REFERENCES AND NOTES

1. G. von Heijne, *J. Mol. Biol.* **184**, 99 (1985).
2. For review, see S. A. Benson, M. N. Hall, T. J. Silhavy, *Annu. Rev. Biochem.* **54**, 101 (1985).
3. M.-J. Gething and J. Sambrook, *Nature (London)* **300**, 598 (1982).
4. K. Sekikawa and C. J. Lai, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3563 (1983).
5. C. A. Kaiser and D. Botstein, *Mol. Cell. Biol.* **6**, 2382 (1986).
6. D. Perlman, P. Raney, H. O. Halvorson, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5033 (1986).
7. C. A. Kaiser and D. Botstein, unpublished results.
8. J. Ngsee and M. Smith, personal communication.
9. D. Perlman, H. O. Halvorson, L. E. Cannon, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 781 (1982).
10. M. Carlson and D. Botstein, *Cell* **28**, 145 (1982).
11. M. Carlson, R. Taussig, S. Kustu, D. Botstein, *Mol. Cell. Biol.* **3**, 439 (1983).
12. The probability (*P*) that an inserted sequence will maintain the correct reading frame and will not lead to chain termination is given by $P = (61/64)^n/3$. Where n is the length of an insert in base triplets. This calculation is based on the assumption that all codons are represented with equal frequency. The mean probability for 22 DNA inserts whose size was determined by restriction enzyme mapping is about 4 percent.
13. The AUG codon for the initiator methionine shown for *suc2-827* is 11 base pairs positioned 3′ of the normal *SUC2* initiator AUG. Since this is the only AUG that is in frame with the *SUC2* coding sequence, it is the only reasonable position for translation to begin. The low level of expression specified by this allele may reflect inefficient translational initiation at an AUG downstream of the normal initiator codon.
14. B. Esmon, P. Novick, R. Schekman, *Cell* **25**, 451 (1981).
15. I. Schauer, S. Emr, C. Gross, R. Schekman, *J. Cell Biol.* **100**, 1664 (1985).
16. R. Haguenauer-Tsapis and A. Hinnen, *Mol. Cell. Biol.* **4**, 2668 (1984).
17. B. Lewin, *Gene Expession 2: Eukaryotic Chromosomes* (Wiley, New York, 1980).
18. S. D. Emr *et al.*, *Mol. Cell. Biol.* **4**, 2347 (1984).
19. G. von Heijne, *Eur. J. Biochem.* **133**, 17 (1983); D. Perlman and H. O. Halvorson, *J. Mol. Biol.* **167**, 391 (1983).
20. A. Goldstein and J. O. Lampen, *Methods Enzymol.* **42**, 504 (1975).
21. A. Wyman and R. White, *Proc. Natl. Acad. Sci. U.S.A.* 77, 6754 (1980).
22. H. Ito, Y. Fukuda, K. Murata, A. Kimura, *J. Bacteriol.* **153**, 163 (1983) as modified by C. Kuo and J. L. Campbell [*Mol. Cell. Biol.* **3**, 1730 (1983)].
23. Individual yeast transformants were grown overnight on solid YEP medium with 2 percent glucose. Cells (about 2×10^7) were scraped from the plates and suspended in 1 ml of liquid YEP medium with 0.1 percent glucose and incubated at 30°C for 3 hours to induce the synthesis of regulated invertase. The cells were washed once in 25 m*M* tris, *p*H 7.5, 10 m*M* sodium azide, centrifuged, and lysed in 20 μl of sample buffer [80 m*M* tris, *p*H 6.8, 2 percent SDS, 2 m*M* PMSF, 0.1*M* dithiothreitol, 10 percent glycerol, 0.01 percent bromphenol blue] by agitation for 1 minute in the presence of 0.1 g of glass beads (0.5 mm). Another portion (30 μl) of buffer was added, the lysate was heated to 90°C for 3 minutes, and the cell debris was removed by centrifugation at 12,000*g* for 1 minute. A portion (5 μl) of the supernatant of each of 407 samples was resolved by SDS-PAGE (*30*), the proteins were electrophoretically transferred to nitrocellulose sheets, and invertase antigen was identified with antiserum to invertase and ^{125}I-labeled protein A (Amersham) (*5*).
24. J. D. Boeke, D. J. Garfinkel, C. A. Styles, G. R. Fink, *Cell* **40**, 491 (1985).
25. O. Gabrial and S.-F. Wang, *Anal. Biochem.* **27**, 545 (1969).
26. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
27. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**, 105 (1982).
28. D. M. Engleman, T. A. Steitz, A. Goldman, *Annu. Rev. Biophys. Chem.* **15**, 321 (1986).
29. The amino acid sequences for the precursor forms of *Saccharomyces cerevisiae* secreted and membrane proteins were obtained from P. L. Liljestrom [*Nucleic Acids Res.* **13**, 7257 (1985)] and J. Thorner (personal communication). The amino acid sequences of cytoplasmic yeast proteins were obtained from the Protein Sequence Database of the Protein Identification Resource supported by the Division of Research Resources of the National Institutes of Health.
30. U. K. Laemmli, *Nature (London)* **227**, 680 (1970).
31. R. B. Trimble and F. Maley, *J. Biol. Chem.* **252**, 4409 (1977).

32. We thank Eric Lander and Mark Daly for assistance in the analysis of protein sequences, Kim Arndt for advice on DNA sequencing, and Arlene Wyman for the human DNA. Supported by PHS grants GM18973 and GM21253 and grant MV90 from the American Cancer Society, a Swanson Fellowship to C.K., and an NSF graduate fellowship to D.P.

20 October 1986; accepted 18 December 1986