The Functional Efficiency of a Mammalian Signal Peptide Is Directly Related to Its Hydrophobicity*

(Received for publication, November 14, 1989)

Phillip Bird, Mary-Jane Gething, and Joe Sambrook
From the Department of Biochemistry and §Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75235

We have previously shown that the signal sequence of the *Saccharomyces cerevisiae* vacuolar protein carboxypeptidase Y (CPY) does not function in mammalian cells unless a glycine residue in the central core is replaced by leucine. Additional mutants were constructed to investigate the features of this hydrophobic core (h) region that are important for signal sequence function in mammalian cells. We find that the degree of hydrophobicity of the h region of any particular mutant signal is directly related to the efficiency with which it directs the translocation of CPY. A minimal h region in a functional signal appears to consist of five hydrophobic residues interrupted by 1 glycine. Analyses of potential secondary structures suggests that a functional mutant signal is more likely than the non-functional CPY signal to adopt either a β strand or an α-helical conformation.

The majority of secreted and membrane-bound proteins from either prokaryotic or eukaryotic organisms possess an amino-terminal signal peptide that is cleaved from the nascent precursor polypeptide during biosynthesis (for reviews, see Rapoport, 1986, and Gierasch, 1989). Analyses of a large number of these signal peptides have revealed a common structural motif that occurs in the absence of significant changes in amino acid sequence homology. In general, a signal sequence appears to consist of a positively charged amino (n) terminus, a hydrophobic core (h) region, and a more polar carboxyl (c) terminus that defines the signal peptidase cleavage site (von Heijne, 1985).

Studies on the structural features of signals have primarily involved bacterial proteins and have suggested that the h region is critical to signal sequence function (Gierasch, 1989). Disruption of the h region by deletion or by replacement of hydrophobic residues with hydrophilic or charged amino acids leads to loss of signal function (Emr and Silhavy, 1983; Bankaitis et al., 1985; Bedoulle et al., 1980), whereas alterations to the n region have little effect (Duffaud et al., 1985). These results suggest that a crucial property of the hydrophobic core is the ability to adopt a periodic structure such as an α helix or β strand (Bedoulle and Hofnung, 1981; Emr and Silhavy, 1983; Kendall et al., 1986). Physicochemical studies on the behavior of synthetic signal peptides in various environments support this suggestion (Briggs and Gierasch, 1984), while other results indicate that it is overall hydrophobicity of the h region which is important (Iida et al., 1985; Ryan and Bassford, 1985). For example, Kaiser et al. (1987) have shown that the signal sequence of *Saccharomyces cerevisiae* invertase can be functionally replaced by randomly generated stretches of predominantly hydrophobic amino acids, while we have shown that the signal sequence of the *S. cerevisiae* vacuolar protein carboxypeptidase Y (CPY) does not function in mammalian cells unless a glycine residue in the h region is replaced with a leucine (Bird et al., 1987).

In *S. cerevisiae*, CPY is translocated, glycosylated, and undergoes signal cleavage (Blachley-Dyson and Stevens, 1987). We have previously demonstrated that CPY expressed in mammalian cells is localized in the cytoplasm, is not glycosylated, and retains its signal peptide (Bird et al., 1987). One explanation for these results is that the CPY signal sequence, which has a core of 5 hydrophobic residues interrupted by 2 glycines, violates von Heijne's (1985) predictions that a minimal h region will consist of 7 hydrophobic residues interrupted by no more than 1 glycine, proline, serine, or threonine. By altering the CPY signal sequence to replace either or both glycines in the h region with leucine, we were able to produce derivatives which function in mammalian cells (Bird et al., 1987). Since it can be converted by a single amino acid substitution from a nonfunctional to a functional signal, we thought that CPY might prove to be a good model with which to further investigate the structural requirements of signal sequences functioning in mammalian cells. We show that a minimal h region may consist of 5 hydrophobic residues interrupted by 1 glycine, that the efficiency of translocation is directly related to the hydrophobicity of the h region, and that a functional signal appears to be more likely to adopt a β strand or an α helical conformation.

**MATERIALS AND METHODS**

*Plasmid Constructions*—The basic strategy employed to construct the mutants was similar to that used in the construction of pSVTCPYm1-m3 (Bird et al., 1987). The starting plasmid pSVTCPY (or pSVTCPYm1 in constructing pSVTCPYm6) was digested with EcoRI and SmaI. Single-stranded DNA of the resulting M13 bacteriophage derivative was prepared using the strain CJ236 (Bio-Rad Laboratories), and the resulting termini were filled using the large (Klenow) fragment of *Escherichia coli* DNA polymerase I. Following further digestion with EcoRI, the (smaller of two) fragment containing the CPY signal sequence was isolated and cloned into M13mp19 replicative form (RF) DNA which had been cleaved with EcoRI and SmaI. The abbreviations used are: CPY, carboxypeptidase Y; RF, replicative form; SDS, sodium dodecyl sulfate; SRP, signal recognition particle.
the CPY signal. The mutagenesis was performed essentially as described by Kunkel (1985) except that T4 DNA polymerase was used instead of the Klenow enzyme. In most cases (7/9), mutants were identified by sequencing DNA from four primary plaques. The remaining mutants were identified by standard screening techniques (Zoller and Smith, 1984). All putative mutants were plaque-purified and resequenced before RP DNA from each was digested with EcoRI and BglII (which cuts upstream of the NcoI site in CPY). The CPY fragment from each mutant was isolated and ligated to an aliquot of the purified large EcoRI-BglII fragment of pSVTCPY, which contains the vector and 3' sequences of CPY. Finally, each of the resulting plasmids, pSVTCPYm4 through m12 was digested with EcoRI and BglII, ligated to M13mp19 RF DNA cut with EcoRI and BamHI, and the region of LNA encoding the signal peptide was resequenced. The deduced amino acid sequences of the mutated signals are shown in Fig. 1.

Transfection, Radiolabeling, and Fractionation of COS-1 Cells—COS-1 cells (Gluzman, 1981) were transfected, radiolabeled, and fractionated as described previously (Bird et al., 1987). Cell lysates were immunoprecipitated using rabbit anti-CPY antiserum obtained from Dr. T. Stevens and were analyzed using 8% SDS polyacrylamide gels and fluorography as previously described (Bird et al., 1987).

Secondary Structure Predictions—Potential secondary structures of the CPY signal sequence and the mutant sequences were predicted using the Chou and Fasman (1978) algorithm through the Intelligent Genetics (Mountain View, CA) BIONET.

RESULTS

Design and Construction of CPY Mutants—To continue our study of the structural requirements of signal sequences functioning in mammalian cells, we designed and constructed a new series of plasmids, pSVTCPYm4-m12, each of which carries a different mutation in the h region of the CPY signal sequence (Fig. 1). These mutants fall into three broad categories designed to test whether the efficiency of a signal peptide is related to the length, hydrophobicity, or number of glycines within its h region.

The first category of mutant was designed to determine whether the CPY signal is nonfunctional in mammalian cells because it contains more than 1 glycine in a minimal h region (von Heijne, 1985). To test this idea, the mutant m12 was designed simply to delete 1 of the 2 glycines (at position 10) in the h region. An alternate proposition is that the CPY h region is not hydrophobic enough (Bird et al., 1987), and that increasing its hydrophobicity without altering the number of glycines would be sufficient to make it functional. The mutants m4 and m5 were therefore constructed by replacing the cysteine at position 9 with alanine or leucine, respectively. These particular residues were chosen because with respect to cysteine they either decrease (alanine) or increase (leucine) the hydrophobicity of the h region.

The second type of mutant is represented by pSVTCPYm6, which was designed to determine whether shortening the length of the h region of a functional signal without removing a hydrophobic residue can influence the efficiency with which it functions. This mutant was derived from pSVTCPYm1 (Bird et al., 1987), in which leucine replaces glycine at position 10, by deleting the remaining h region glycine at position 12.

The third and final category of mutant was designed to determine whether varying the overall hydrophobicity of the CPY h region results in differences in translocation efficiencies. From our previous work, we know that a substitution of glycine for leucine at position 10 in the CPY h region results in a signal that functions well in mammalian cells (pSVTCPYm1; Bird et al., 1987). Because leucine is one of the most hydrophobic amino acids, we were interested to see whether substitution with a less hydrophobic residue would also enable the CPY signal to function and, if so, whether it would result in a signal peptide of differing functional efficiency to CPYm1. Therefore, in each of the mutants pSVTCPYm7-11, the glycine at position 10 is replaced by a different hydrophobic amino acid.

Expression of the CPY Mutants in COS-1 Cells—To assay these mutations in the CPY signal sequence, we transfected COS-1 cells with plasmids carrying either the wild type or mutant versions of the CPY gene. CPY production was demonstrated by radiolabeling transfected cells using [35S]methionine, immunoprecipitating CPY from cell extracts, and analyzing it by SDS polyacrylamide gel electrophoresis. From our previous studies, we know that after a 30–60 min labeling wild type CPY is detected as a 62-kDa protein in the cytoplasm of transfected cells. By contrast, glycosylated mutants of CPY produced in pulse-labeled COS-1 cells typically exhibit a size distribution of 68–75 kDa, which consists of a mixture of core-glycosylated, partially processed, and terminally glycosylated CPY (Bird et al., 1987). If labeled in the presence of tunicamycin, an antibiotic which inhibits N-linked glycosylation (Tkacz and Lampen, 1975), there is an apparent decrease in the size of the molecule (to about 61 kDa) due to the lack of carbohydrate side chains on the nascent polypeptide. The slightly smaller size of the polypeptide chain observed in these circumstances (61 kDa) compared to that of nontranslocated CPY (62 kDa) is presumably due to the removal of the 25-residue signal peptide from translocated molecules (Bird et al., 1987).

The plasmids pSVTCPYm1-m12, together with pSVTCPY and pSVTCPYm1 as controls, were used to transfect COS-1 cells which were subsequently radiolabeled in the absence or presence of tunicamycin and analyzed as described above. All the new CPY mutants were efficiently expressed in COS-1 cells (Fig. 2). Interestingly, in the absence of tunicamycin, we observed (with the exception of CPYm4) that each mutant protein was produced in both the 62-kDa and the 68–75-kDa forms, while in the presence of tunicamycin there was only a 61–62-kDa form apparent (Fig. 2). In addition, it appeared that the ratio of the 62-kDa to 68–75-kDa forms was different in each case; ranging from CPYm4 and CPYm6 which were observed predominantly in the 62-kDa or in the 68–75-kDa forms, respectively, to CPYm5 which was equally distributed.

The simplest explanation for these observations is that most of these new mutations in the CPY signal enable it to function only partially in mammalian cells. In the majority of cases, this results in a pool of cytoplasmic, unglycosylated (62-kDa) proteins and a population of translocated and glycosylated (68–75-kDa) molecules. Densitometric scans of gels such as those represented in Fig. 2 showed that in each case the two forms of CPY accounted for 70 ± 10% of the total radioactivity present in the immunoprecipitate (data not shown). Coupled with the fact that the half-life of CPY in the
cytoplasm is greater than 1 h following a 15-min pulse-labeling, this suggested that during the period of labeling there was negligible turnover of either the cytoplasmic or the translocated forms of the protein. Thus, to estimate the efficiency with which each particular mutant was translocated, at least three gels similar to the one shown in Fig. 2 were scanned, and the relative proportion of glycosylated to nonglycosylated CPY was calculated and expressed as the average percentage translocated (Fig. 1).

**Fractionation of COS-1 Cells Transfected with the CPY Mutants**—To confirm that partial translocation of the mutants was indeed occurring, we decided to examine the subcellular distribution of the forms of CPY that were observed in each case. We have previously shown that COS-1 cells expressing CPY can be fractionated into cytosolic (C) and microsomal (M) fractions and that CPY is located in the cytosolic fraction whereas a translocated protein is present predominantly in the microsomal fraction (Bird et al., 1987).

Radiolabeled COS-1 cells expressing the mutants were gently homogenized, the nuclei and large cellular debris were removed, and then the resulting lysate was separated into cytosolic and microsomal fractions by ultracentrifugation. CPY was immunoprecipitated from each fraction and analyzed by SDS-polyacrylamide gel electrophoresis and fluorography (Fig. 3). Under these experimental conditions, there is some leakage of translocated proteins (including 68-75-kDa CPY) into the cytosolic fraction, but not vice versa (Bird et al., 1987).

Although a variable but significant amount of background was observed in immunoprecipitates from the microsomal fractions, it was clear that forms of most mutants were present predominantly in the microsomal fraction (Bird et al., 1987). These results confirm that a range of mutations in the CPY signal sequence enable it to function in mammalian cells but demonstrate that most are only partially effective.

**Hydropobicity and Potential Conformations of the Mutant CPY Signals**—It has been repeatedly suggested that the hydrophobic core of a signal sequence is crucial for its function (see Rapoport, 1986). However, a successful analysis of the effects of substituting residues in the h region with amino acids of varying hydrophobicity depends on the availability and reliability of methods for ranking amino acids in order of hydrophobicity (for review, see Rose et al., 1985a). To calculate and compare the hydrophobicities of the CPY signal sequence and its derivatives, we used three scales. The first (Levitt, 1976; modified by Hopp and Woods, 1981) is a series of solvent parameters derived from measurements of the free energy required to transfer amino acid solutes from an aqueous to an organic phase (Nozaki and Tanford, 1971). The second (Rose et al., 1985b) is derived from an analysis of proteins of known structure which measures the average area that each amino acid buries in the interior of a folded protein. This average area buried correlates with residue hydrophobicity. The third scale (Sweet and Eisenberg, 1983) is an empirical ranking that was derived by an iterative procedure requiring an initial scale of residue hydrophobicities together with the observed frequencies of amino acid substitutions within related proteins. Upon averaging the initial hydrophobicities over residue replacements from related structures, new values were evolved which are largely independent of the initial ones. In all three scales, a numerical value is assigned to each amino acid and the more hydrophobic the residue, the higher the value.

Considering the 7-residue h region of CPY and the signal mutants (6-residue regions in the case of the deleted variants CPYM6 and m12), we used these scales to calculate the average hydrophobicity of each particular region. This was derived by adding together the assigned values of the residues comprising a particular h region then dividing by the number of residues making up that region (Fig. 4). Of the three scales used, we found that the Levitt and Sweet rankings enabled us to establish a reasonable correlation between hydrophobicity and translocation efficiency: the more hydrophobic the h region, the more efficient the signal. The results obtained

---

2 P. Bird, M.-J. Gething, and J. Sambrook, unpublished results.
using the Rose scale are generally consistent with this correlation, however, since cysteine is ranked higher than leucine on this scale, the partially functional mutant CPYm5 is calculated to be less hydrophobic than CPY, while CPYM11 which translocates at 44% is calculated to be more hydrophobic than other mutants that translocate at 90% or better.

In addition to the contribution of hydrophobicity to signal sequence function, it has been suggested that a periodic secondary structure such as an α helix or β strand, respectively important (Gierasch, 1989). We used the Chou and Fasman (1978) algorithm to predict secondary structures that are likely to be adopted by the CPY signal sequence and its mutants (Fig. 5). Although it was developed to predict structures in globular proteins, this algorithm has been used previously to predict structures in the lamB and preproparathyroid hormone signals which have been subsequently confirmed by physiochemical studies using synthetic peptides (Emr and Silhavy, 1983; Briggs et al., 1986; Rosenblatt et al., 1980). We find that the CPY signal is predicted to adopt either a β strand or α helix but is unstructured around residues 10–12 in the h region. By contrast, most of the mutations which yield functional signals serve to increase the likelihood of β strand or α helix formation both in this and the flanking regions. There also seems to be a correlation between the degree of function and the probability of forming a fully structured h region. For example, the most efficient signals (exemplified by CPYM2, -m3, and -m6) which translocate at 92% or better appear to have a higher probability of adopting an uninterrupted β strand.

**DISCUSSION**

Partial Translocation—The first models describing the early stages in translocation postulated a tight coupling between translation and translocation, with the signal recognition particle (SRP) binding to a signal sequence and arresting translation until the ribosome-mRNA-signal-SRP complex interacts with the membrane of the endoplasmic reticulum (for review, see Walter et al., 1984). Although this complex is essential for translocation, translation and translocation are not necessarily mechanistically coupled in that a polypeptide may be almost completely synthesized before translocation commences (Meyer, 1985; Perera et al., 1986; Mueckler and Lodish, 1986; Chao et al., 1987). In some cases, it appears that SRP may bind to the signal sequence after most of the polypeptide has been translated (Ainger and Meyer, 1986), while in others a successful SRP-signal interaction only occurs early in translation (Rothblatt and Meyer, 1986). To account for these apparently contradictory findings, it has been suggested that SRP has the potential to bind to a signal at any time during translation but may be limited to a particular temporal window because the signal is eventually sequenced as the protein folds. Thus, proteins which are apparently post-translationally translocated may have signals which are accessible to SRP after translation is virtually complete, while those that are not post-translationally translocated fold in such a manner that the signal is inaccessible to SRP.

Expanding this model further, it has been proposed that any SRP-signal interaction is in equilibrium between bound and nonbound and has a measurable affinity constant (Raaport et al., 1987). This binding affinity may vary from signal sequence to signal sequence, so it is envisaged that proteins which tend to fold into translocation-incompetent forms would have signals of higher affinity for SRP to increase the likelihood of a productive interaction early in translation. Our finding that most of the CPY mutant signals are partially functional and that there is a range of translocation efficiencies is consistent with this “equilibrium” model. We suggest that each signal has a different affinity for SRP, ranging from high (m1-m3, -m6, -m7, -m9, -m10) to medium (m5, m11) and low (m8, m12) affinities. Thus, the nontranslocated, cytoplasmic forms of CPY that we observe probably represent polypeptides which are synthesized without the establishment of a successful signal-SRP interaction. Interestingly, measurement of the affinities of SRP binding to the CPY and m1-m3 signals in vitro shows that the mutant signals have higher affinities than the normal CPY signal.³

³ W. Hansen, personal communication.

**Hydrophobicity and Signal Sequence Function**—Despite observations which correlate the hydrophobicity of the h region with signal sequence function, no systematic study of this relationship has been performed. This may be due in part to the difficulty in establishing a reliable ranking of the amino acids with respect to hydrophobicity (for a discussion, see Rose et al., 1985a), or it may be due to a failure to isolate

![Table](http://www.jbc.org/)

**Fig. 5. Potential conformations of the CPY signal peptides.** (Px) = the average propensity for x = α(pha helix) or β(eta sheet). A and B designate α helix and β strand, respectively. Gaps indicate regions of no predicted secondary structure (Chou and Fasman, 1978).
many mutants which retain partial signal function. One problem in identifying such mutants is that untranslocated, cytoplasmic derivatives of normally secreted proteins are sometimes unstable (Gething and Sambrook, 1982).

In the mutants CPY m1 and m7-m11, the glycine at position 10 in the CPY signal has been replaced by a more hydrophobic amino acid, and all are at least partially functional. Use of three independently derived scales to calculate hydrophobicities shows that all of these mutants possess h regions of higher overall hydrophobicity than CPY and suggests that the higher the hydrophobicity the more efficient the signal. Indeed, when the Levitt or Sweet scales are used to calculate the average residue hydrophobicity, a linear relationship between hydrophobicity and the degree of translocation becomes apparent between a lower limit represented by CPY and an upper limit represented by CPY m7 (those translocating at 90% or better). Only m9 does not completely fit this trend, in that it is calculated to be more hydrophobic than m7 but translocates at a slightly lower efficiency. This may reflect uncertainty in the ranking of phenylalanine or a slightly average hydrophobicity per residue is greater. This suggests that the average hydrophobicity per residue must be greater than a threshold value to enable a signal to function.

The concept of hydrophobicity being important to signal function is also supported by the behavior of the mutants m4 and m5. The m4 mutant is not detectably functional, and on all scales is calculated to have an overall hydrophobicity lower than CPY itself because alanine is considered to be less hydrophobic than cysteine. On the other hand, m5 is 43% functional, and using the Levitt and Sweet scales can be shown to be more hydrophobic than CPY. (The Rose scale ranks cysteine above leucine, which accounts for the lower hydrophobicity of m5 compared to CPY calculated using this scale.)

Finally, the mutant m12 has the glycine residue at position 10 deleted, resulting in a 6-amino acid h region. This derivative is partially functional even though it consists of only 5 hydrophobic residues plus 1 glycine. Calculation of the hydrophobicity of the m12 h region using the three scales shows that although it is similar in total to the CPY h region, the average hydrophobicity per residue is greater. This suggests that the average hydrophobicity per residue must be greater than a threshold value to enable a signal to function.

Length and Composition of a Minimal h Region—von Heijne (1985) has analyzed a large number of signal sequences and demonstrated that much of the variation in the length of signal peptides occurs in their hydrophobic cores. These h regions are also frequently punctuated by the small polar residues glycine, proline, threonine, or serine. In attempting to define the limits of variation among signal sequences, von Heijne (1985) has also suggested that a minimal h region consists of 6 hydrophobic residues interrupted by no more than one of the small polar amino acids. An example of a signal possessing such a minimal h region is the chicken α2 [I]-collagen peptide, which consists of seven hydrophobic amino acids with no interruptions (von Heijne, 1985).

By contrast, our mutants m6 and m12 have shorter h regions consisting of 6 residues and therefore allow a new lower limit for the length of an h region to be defined. Although they are the same length, these mutant signal peptides exhibit very different functional efficiencies. An obvious difference between the two is that m12 contains an interrupting glycine which in m6 is substituted by a leucine. The fact that m6 functions more efficiently than m12 is therefore consistent with von Heijne's argument that a minimal h region is unlikely to contain more than 1 small polar residue and shows that a signal with a noninterrupted minimal h region may function more efficiently than an interrupted one.

The underlying reason for such limits to signal sequence variation may lie in the overall hydrophobicity rather than length of the h region. For example, m12 was generated by deleting a glycine from the CPY signal sequence resulting in a partially functional signal with a smaller, more hydrophobic h region. Likewise, m6 was generated in a similar fashion from m1, yielding a shorter signal with a more hydrophobic h region and increased functional efficiency. These findings support the concept that limits on hydrophobicity (Iida et al., 1985; Ryan and Bassford, 1985) rather than length (Bedoulle and Hofnung, 1981) define a minimal h region.

Secondary Structure and Signal Sequence Function—Analyses of periodic structures within proteins reveal that hydrophobic residues are usually found in α helices or β strands (Chothia, 1984). Previous studies on signal sequence function have therefore considered the possible relationship between secondary structure and function (Gierasch, 1989). Studies on the bacterial λamB signal, its mutants, and revertants using the Chou-Fasman algorithm suggest that functional signals adopt an α-helical structure (Emr and Silhavy, 1983). This is supported by work on synthetic λamB or derivative signal peptides which show that those demonstrated to function in vivo tend to adopt an α helical conformation in micellar solutions (Briggs and Gierasch, 1984; Briggs et al., 1985), are unstructured in aqueous solutions but adopt a β structure on electrostatic interaction with a phospholipid monolayer (Briggs et al., 1986). Similar studies on a synthetic mammalian signal peptide derived from preproparathyroid hormone suggest that the signal may adopt either an α-helical or β strand conformation, depending on its environment (Rosenblatt et al., 1980).

We find that our functional mutant signals are more likely to have structured h regions than CPY, with a β strand conformation being slightly favored over α helix. This is reminiscent of both the λamB and the preproparathyroid hormone signals; however, it will be necessary to examine the behavior of synthetic CPY signal peptides in various environments before making a more conclusive statement about the relationship between conformation and function.

Acknowledgments—We are grateful to Tom Stevens (University of Oregon, Eugene, OR) for providing the CPY antisera. We also thank Mark Segal for help with the computer analysis, Jamie McNight and Lila Gierasch for useful discussions, and Sue Ying Huang for technical assistance.

REFERENCES


80, 4599–4603
78, 3824–3828
Iida, A., Groarke, J. M., Park, S., Thom, J., Zabicky, J. H., Hazelbauer, 
**235**, 312–317
706–708
348–352
Chem._ **37**, 1–109
Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, 
M. H. (1985b) _Science_ **229**, 834–838
14837
Commun._ **65**, 248–257
The functional efficiency of a mammalian signal peptide is directly related to its hydrophobicity.

P Bird, M J Gething and J Sambrook