

Single Amino Acid Substitutions Can Convert the Uncleaved Signal-Anchor of Sucrase-Isomaltase to a Cleaved Signal Sequence*

(Received for publication, April 14, 1992)

Martin Hegner, Annette von Kieckebusch-Gück, Rocco Falchetto, Peter James, Giorgio Semenza, and Ned Mantei†

From the Department of Biochemistry, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

A hydrophobic segment near the amino terminus (positions 12–32) of rabbit sucrase-isomaltase functions both as a membrane anchor and as a signal sequence for translocation into the endoplasmic reticulum. Unlike most signal sequences, that of sucrase-isomaltase is not cleaved by signal peptidase. Using *in vitro* transcription and translation systems, we have found that substitution of a single proline, at position 28 or 29, converted the signal-anchor to a cleaved signal sequence, with cleavage occurring after alanine 26 and the introduced proline thereby occupying position +2 or +3 relative to the cleavage site. Two deletions that shorten the transmembrane domain by 8 amino acids were also effective, whereas various other changes upstream and downstream of this domain were without effect. We conclude that susceptibility to mammalian signal peptidase is influenced both by the length of the hydrophobic region and by the secondary structure downstream of the cleavage site.

The rabbit sucrase-isomaltase (SI)¹ complex of the brush border membrane is synthesized as a single-chain precursor of 1827 amino acids that is then split by pancreatic proteases to form the sucrase and isomaltase subunits. SI is a member of a major group of stalked brush border proteins that are anchored via a hydrophobic segment located not far from the amino terminus (monotopic proteins of type II). In the case of SI, the transmembrane domain lies at the N terminus of the isomaltase subunit. This domain functions both as a permanent membrane anchor and as a signal peptide that directs targeting to the endoplasmic reticulum (ER). In contrast, monotopic proteins anchored at the carboxyl-terminal region (*i.e.* type I), secretory proteins, and proteins anchored via glycosylphosphatidylinositol possess as a rule an amino-terminal signal peptide that exists only transiently, because it is removed through the action of the signal peptidase in the ER.

von Heijne (1983, 1984a, 1984b, 1985, 1986a, 1986b) has tabulated many of the known *cleaved* signal sequences and attempted to distill out possible rules or regularities in their

structure. The sequences range in length from 13 to 36 amino acids. A cleaved signal sequence can typically be divided into three subregions (von Heijne, 1985): 1) a polar, usually positively charged, amino-terminal part (n region), 2) a hydrophobic core, usually 8–12 amino acids long (h region), and 3) a usually more polar stretch of 4–7 amino acids between the h region and the cleavage site (c region). There is very little sequence homology among signal sequences, so that a particular sequence almost certainly plays no decisive role (Briggs and Gierasch, 1986).

The sequence near the amino terminus of rabbit prosucrase-isomaltase (pro-SI) (Hunziker *et al.*, 1986) is shown in Fig. 1A. An n region of 11 amino acids (not including the initiator methionine) with a net charge of +3 is followed by an h region 20 residues long. This sequence, or parts of it, serves both as signal and as membrane anchor (with the amino terminus inside and the carboxyl terminus outside) and is not cleaved by signal peptidase (Ghersa *et al.*, 1986; Hunziker *et al.*, 1986).

What structural elements might differentiate noncleaved signal sequences, such as that of pro-SI, from cleaved signal sequences? The length of the n region of pro-SI, 11 amino acids, is longer than average but not extraordinary. Indeed, the effect of n region length is not consistent. Uncleaved signal sequences may have very short n regions (*e.g.* 6 amino acids for influenza virus neuraminidase (Bos *et al.*, 1984)), whereas moving a normally cleaved signal sequence to the interior of a protein does not necessarily prevent cleavage (Perara and Lingappa, 1985; Beltzer *et al.*, 1989). On the other hand, drastic shortening of a long n region can lead to cleavage of a previously uncleaved signal-anchor (Lipp and Dobberstein, 1986; Schmid and Spiess, 1988).

The clearest difference between cleaved signals and uncleaved amino-terminal signal/membrane anchors is in the length of the h region, averaging 8–12 residues for the former and about 20 for the latter (von Heijne, 1985; Briggs and Gierasch, 1986). However, as mentioned above, long h regions sometimes become cleavable when their n regions are shortened.

The segment between the h region and the cleavage site is defined as the c region. Based on a statistical analysis, von Heijne (1983) proposed a “(–3, –1) rule”: at –1 from the cleavage site (cleavage is defined as occurring between –1 and +1), there is a small, neutral amino acid (Ala, Ser, Gly, Cys, Thr, or Gln); position –3 may not be aromatic, charged, or large polar. The former requirements have been in general confirmed by construction and analysis of an extensive series of mutations in position –1 of the pre(Δpro)apolipoprotein A-II signal sequence (Folz *et al.*, 1988). An amino acid with α-helix-breaking properties, such as proline or glycine, may be important near the start of the c region (Vlasuk *et al.*, 1984; Yamamoto *et al.*, 1989). For example, some of the poorly processed mutations at position –1 in the

* Supported by the Swiss National Science Foundation, Berne. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed. Tel.: 41-1-256-3050; Fax: 41-1-252-8744.

¹ The abbreviations used are: SI, sucrase-isomaltase; endo F, endoglycosidase F/N-glycosidase F; pro-SIΔ, a plasmid coding the 230 amino-terminal amino acids of SI; SDS, sodium dodecyl sulfate; ER, endoplasmic reticulum; Tricine, N-tris(hydroxymethyl)methylglycine.

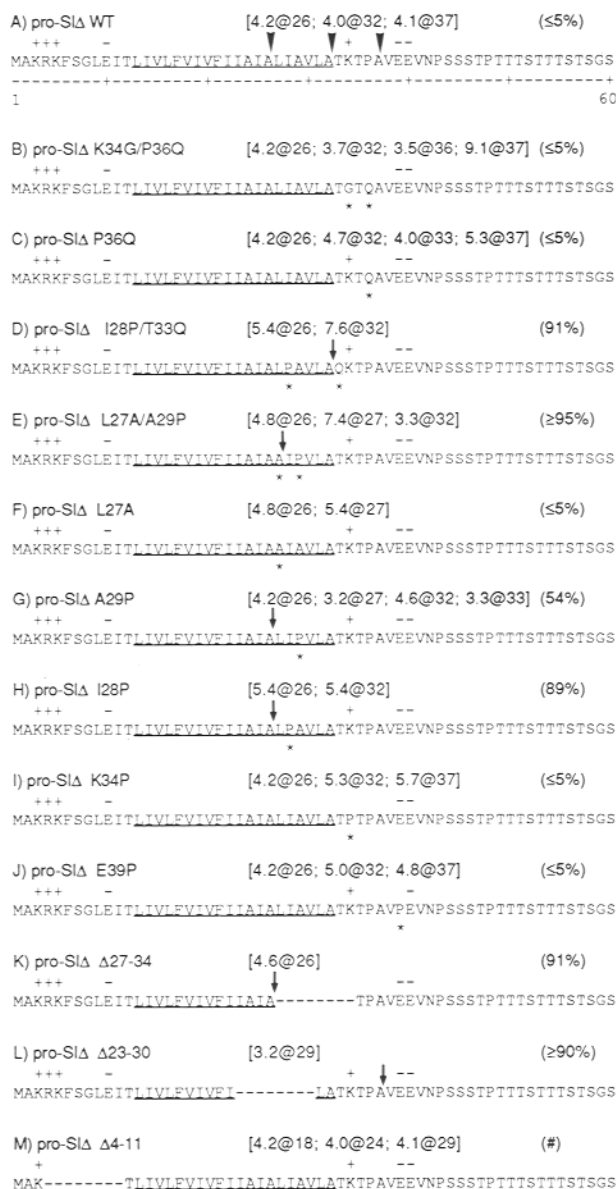


FIG. 1. Sequence of the N-terminal proximal region of rabbit sucrase-isomaltase coded by pro-SIA wild type and derivatives. Charged amino acids are indicated by plus and minus signs; the hydrophobic transmembrane region is underlined; amino acids changed in mutants are indicated with asterisks; deletions are indicated by dashed lines. Vertical arrows indicate the cleavage sites observed; the arrowheads in A indicate potential (but not actually cleaved) sites. Numbers in square brackets indicate cleavage scores calculated for the various positions, using the algorithm of von Heijne (1986a); for example, 4.2@26 indicates a score of 4.2 for cleavage after residue 26. Numbers in parentheses indicate the per cent cleavage of translocated product by signal peptidase. Less than ~5% cleaved or uncleaved form could not be reliably distinguished from the background (cf. Fig. 2). #, not translocated, so that per cent cleaved was not applicable (see "Results").

pre(Δ pro)apolipoprotein A-II signal (Folz *et al.*, 1988) could be partly rescued by introducing a proline at position -5 (Nothwehr and Gordon, 1989).

Downstream sequences have been shown to play a role in determining the exact cleavage site. An effect of the residue at +1 has been demonstrated by analysis of a panel of mutants (Nothwehr *et al.*, 1990). Deletions and other mutations in mature domains can affect either signal sequence cleavage (Folz and Gordon, 1986; Burgess *et al.*, 1987) or translocation efficiency (Folz and Gordon, 1987; Andrews *et al.*, 1988).

These appear to be at least in part global effects, depending on more than just the immediate neighbors of the c region, as internal deletions of the mature domain can affect translocation (Andrews *et al.*, 1988).

Finally, we consider a weight matrix method for predicting signal sequence cleavage sites, which considers contributions from part or all of the h region, the c region, and the following 2 residues (*i.e.* from position -13 to +2) (von Heijne, 1986a). Each amino acid within this window is given a weight related to its frequency of occurrence at this position in signal sequences, as compared with its frequency in proteins in general. Alanine at position -1, for example, has a high positive weight; proline is negatively valued at positions -3 to +1 but has a positive weight at the start of the c region in positions -4 and -5. The higher the score for the window as a whole, the more likely the window defines a signal cleavage site. Typical scores for known cleavage sites are in the range of 6-12, with less than 2% of the cleavage sites having a score less than 3.5. Conversely, in a sample of amino-terminal regions of known cytosolic proteins, only 2% had maximal scores greater than 3.5. The algorithm can predict about 80% of known cleaved signals. When such a window was moved along the pro-SI amino terminus, several potential cleavage sites were found, albeit with only borderline scores: 4.2 after Ala-26, 4.0 after Ala-32, and 4.1 after Ala-37.

In sum, if the membrane anchor of pro-SI was once a cleaved signal, the above considerations suggest ways we may be able to induce sensitivity to the signal peptidase, which in turn may provide information as to what the peptidase expects of its substrates. The question of what characteristics prevent permanent, amino-terminal membrane anchors from being split by signal peptidase was the starting point of this work.

EXPERIMENTAL PROCEDURES

Materials—T7 RNA polymerase was purchased from New England Biolabs. Endo F was purchased from Boehringer Mannheim. Other modifying or restriction enzymes and ribonuclease inhibitor were purchased from Pharmacia LKB Biotechnology Inc. and New England Biolabs. Taq polymerase was obtained from Perkin-Elmer Cetus Instruments. L-[U-¹⁴C]Alanine (>150 mCi/mmol) and L-[U-¹⁴C]proline (>290 mCi/mmol) were from Amersham Corp. L-[³⁵S]Cysteine (1100 mCi/mmol), L-[3,4,5-³H]leucine (156 mCi/mmol), and L-[U-¹⁴C]valine (>250 mCi/mmol) were from Du Pont-New England Nuclear. Tetracaine and trypsin were from Sigma, chicken ovomucoid trypsin inhibitor was from Serva Heidelberg, and Nikkol (octaethyleneglycol mono-n-dodecyl ether) was from Fluka Chemicals, Buchs, Switzerland.

cDNA Constructs—All plasmid constructs were made in the vector pBluescriptKS(-) (Stratagene). Plasmid pro-SIA carries a segment of wild-type pro-SI cDNA (Hunziker *et al.*, 1986) that had been originally subcloned into pGem-4, from a BamHI site (18 base pairs upstream of the initiation codon) to a BglII site 701 base pairs downstream. A sequence derived from the original pGem-4 polylinker region another 8 base pairs downstream from the BglII site introduces an artificial stop codon just before a SalI site at a position 713 base pairs downstream. Mutations were introduced into this DNA using the "gapped-duplex" method of Kramer *et al.* (Kramer *et al.*, 1984, 1988; Kramer and Fritz, 1987) or of Kunkel (Kunkel, 1985; Kunkel *et al.*, 1987) or a method based on the polymerase chain reaction (Nelson and Long, 1989). All sequences involved in the mutagenesis experiments were checked by DNA sequencing.

In Vitro Transcription and Translation—Plasmid DNA linearized with SalI was transcribed with T7 RNA polymerase in the presence of 0.5 mM m⁷GpppG (CAP analog) as described by Melton (1987).

Wheat germ cell-free extract for *in vitro* translation was prepared following Anderson *et al.* (1983). Dog pancreas microsomes were isolated and centrifuged according to Walter and Blobel (1983) and Siegel and Walter (1985).

The translations were performed as described by Anderson *et al.* (1983) and Jagus (1987), except that the reaction mixtures were adjusted to final concentrations of 40 mM KOAc and 1.9 mM Mg(OAc)₂, which considerably improved the translation efficiency.

Approximately 400 ng of RNA and either 5 μ Ci of L-[35 S]cysteine, 0.8 μ Ci of L-[3 H]leucine, or 0.4–3.2 μ Ci of one L- 14 C-labeled amino acid were used in a standard 25- μ l reaction. The nonionic detergent Nikkol was included at a final concentration of 0.002% to stabilize signal recognition particle (SRP) activity. In one experiment, a mixture of five protease inhibitors was included: pepstatin, chymostatin, antipain, leupeptin, aprotinin, phenylmethanesulfonyl fluoride, L-1-chloro-3-[4-tosylamido]-7-amino-2-heptanone HCl, and chicken ovomucoid trypsin inhibitor (Mumford *et al.*, 1981; Walter *et al.*, 1981). The mRNA and the membranous fractions (1 eq/25 μ l translation mix) (Walter and Blobel, 1983) were always added last, after all of the other components were mixed and the detergent uniformly diluted. Samples were incubated for 90–120 min at 26 $^{\circ}$ C. To assay for cotranslational translocation, upon completion of translation, samples were adjusted to 3 mM tetracaine HCl (Cioffi *et al.*, 1989) and 150 μ g/ml trypsin and incubated for 20 min at 26 $^{\circ}$ C. In some cases, Triton X-100 was added to 0.8% before trypsin digestion. Trypsin was inactivated by addition of ovomucoid trypsin inhibitor to a final concentration of 150 μ g/ml. Samples to be digested with endo F were adjusted to 0.2% SDS, boiled for 3 min, mixed with 75 μ l of 50 mM potassium phosphate, pH 7.0, 25 mM EDTA, 0.5% *n*-octyl glucoside, 1.3% β -mercaptoethanol and then incubated with 50 milliunits of endo F overnight at 37 $^{\circ}$ C.

All the samples were incubated for 5 min with 0.1 M iodoacetamide to reduce background during gel electrophoresis. After precipitation with methanol/chloroform (2:1) (Wessel and Flügge, 1984), the final samples were boiled in SDS sample buffer (4% SDS, 12% glycerol (w/v), 50 mM Tris-HCl, pH 6.8, 2% dithiothreitol, 0.02% blue G) and analyzed by electrophoresis through 15% SDS-polyacrylamide gels in a Tris glycine buffer system (Laemmli, 1970) or SDS-polyacrylamide gels (9.7% acrylamide, 0.3% bisacrylamide) in a Tricine buffer system (Schägger and von Jagow, 1987). The gels were fixed, soaked in Amplify* (Amersham Corp.), dried, and subjected to autoradiography using Kodak X-Omat AR film for 1–2 days at -80° C. Densitometric scanning was carried out with a Desaga model CD50 chromatogram scanner.

Sequence Determination—For sequence determination, translation was performed with a radioactive amino acid using wheat germ extract as described above, with or without microsomes as appropriate, for 90 min at 26 $^{\circ}$ C using ~ 5 μ g of RNA in a 250- μ l reaction. After trypsin digestion (where appropriate) the products were separated by Tricine-SDS gel electrophoresis, and the gels were electroblotted onto an Immobilon* P membrane. After autoradiography, regions of interest were excised and subjected to automated Edman analysis using an Applied Biosystems 470A protein sequencer. The anilinothiazolinone derivatives were collected directly from the reaction cartridge, and the radioactivity was measured by liquid scintillation counting.

Computer Analysis of Potential Cleavage Sites—The weight matrix method described by von Heijne (1986a) was implemented in a program² written in C for the Macintosh computer.

RESULTS

Experimental Design and Analysis of the Signal-Anchor Sequence of Prosucrase-Isomaltase—We asked what changes in the uncleaved signal-anchor of pro-SI could convert it to a cleaved signal. This question was approached in an *in vitro* system, in which cloned pro-SI cDNA and its variants were transcribed using a bacteriophage RNA polymerase and the resulting RNA translated in a wheat germ cell-free translation system, with or without a mammalian microsomal fraction to allow translocation.

To test whether our mutant signal sequence was cleaved, the complete pro-SI polypeptide (*M*_r 203,000) was clearly too large to be useful. Therefore, a 713-base pair segment of the pro-SI cDNA coding for the 230 amino-terminal amino acids was subcloned into the pBluescript transcription vector, giving plasmid pro-SI Δ . A peptide of molecular weight 27,000 was expected so that the removal of some 20–30 residues by signal peptidase would be apparent. This peptide contains a site for N-linked glycosylation; modification at this site in the lumen of the ER leads to an increase in molecular weight of the product, diagnostic for translocation through the ER

membrane. Cleavage of the signal sequence can be analyzed after removal of the added sugars by treatment with endo F.

Fig. 2A shows the analysis of the *in vitro* translation/translocation of pro-SI Δ wild-type mRNA in a wheat germ system with or without dog pancreas microsomes. Unexpectedly, translation *without* microsomes (lane 1) led to two products, of apparent molecular weight 27,000 and 29,500. In the presence of microsomes, a new band of lower mobility appears (lane 2) of apparent molecular weight 33,000. This band is unaffected by trypsin digestion (lane 3) unless the microsomes are solubilized with Triton X-100 before digestion (lane 5), suggesting the corresponding polypeptide has been translocated into the lumen of the ER. Consistent with this, digestion with trypsin plus Endo H leads to an increase in mobility (lane 4), the apparent weight of the product corresponding to the smaller of two products seen in lane 1. Fig. 2B shows electrophoresis of the same materials using the Tris glycine gel system of Laemmli (1970); the translocated product does not give a band distinguishable from those seen without microsomes. However, it was only with this gel system that cleavage of the signal sequence reliably led to an increase in electrophoretic mobility (compare panels C and D of Fig. 2, representative of the analysis of a cleaved mutant). Highly hydrophobic stretches of amino acids, such as found in the membrane anchor of pro-SI and other signal sequences, are known to occasionally exhibit anomalous mobility in SDS-gel electrophoresis (Waters *et al.*, 1988).

To further characterize the two pro-SI Δ products seen without microsomes and that found after trypsin digestion, [3 H]leucine-labeled translation products, either unmodified or translocated and trypsin-treated, were separated by SDS-gel electrophoresis and radiosequenced as described under "Experimental Procedures." Both polypeptides produced in the absence of microsomes gave similar patterns of release of 3 H, with main peaks consistent with the sequence beginning at Ala-2 (Fig. 3, A and B, vertical arrows). Lesser peaks (Fig. 3, A and B, arrowheads) may indicate removal of 2 further amino acids, perhaps by proteolysis in the wheat germ extract (Mumford *et al.*, 1981). (We expect the series of 3 basic amino acids near the amino terminus to be very sensitive to mono- and dibasic proteases.) However, translation in the presence of a mixture of five protease inhibitors did not affect the

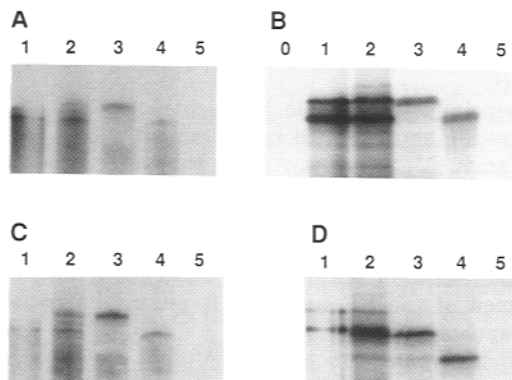


FIG. 2. *In vitro* translation of pro-SI Δ transcript. Translation was carried out in the presence of [35 S]cysteine, in the absence (lane 1) or presence (lanes 0 and 2–5) of dog pancreas microsomes. Aliquots were treated with trypsin (lanes 3), trypsin plus endo F (lanes 4), or trypsin in the presence of Triton X-100 (lane 5). Lane 0, no pro-SI Δ RNA. Lanes 3 and 4 are derived from 25 μ l of translation reaction, the other lanes from 12.5 μ l. The products were separated by SDS-polyacrylamide gel electrophoresis in a Tricine (A and C) or Tris-glycine (B and D) buffer system and visualized by autoradiography. A and B, pro-SI Δ wild-type RNA; C and D, pro-SI Δ L27A/A29P RNA.

² N. Mantei, unpublished program.

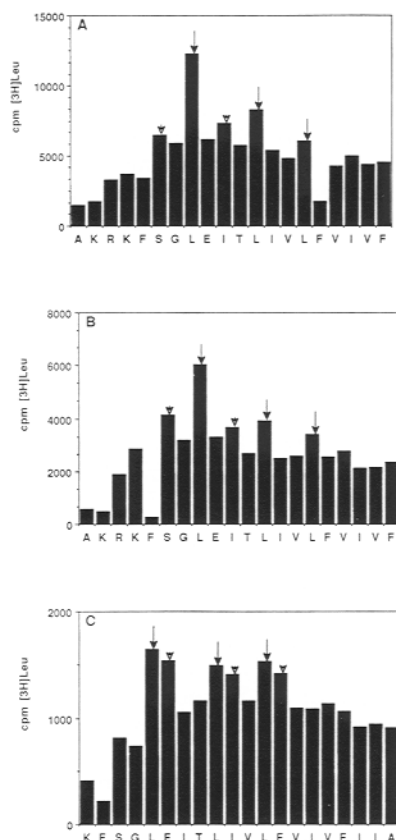


FIG. 3. Radiosequencing of larger (A) and smaller (B) products arising upon translation of pro-SI Δ RNA (cf. Fig. 2A, lane 1) and of the translocated, trypsin-digested product (cf. Fig. 2A, lane 3). [3 H]Leucine-labeled polypeptides were separated by gel electrophoresis and blotted, and regions of interest were subjected to automated Edman degradation. Radioactivity released at each cycle is shown. Arrows in A and B correspond to the peaks expected of a polypeptide beginning at Ala-2 (sequence shown on the abscissa); arrowheads correspond to the peaks expected if the polypeptide begins with Arg-4. In C, arrows correspond to products beginning with Lys-5 (abscissa); arrowheads correspond to those beginning with Arg-4.

smaller product. Translation at a higher temperature (30 °C) or in the presence of methyl mercury hydroxide (Payvar and Schimke, 1979), conditions which can meliorate inhibitory effects of mRNA secondary structure on translation, also had no effect.

Sequencing of the translocated peptide isolated after trypsin digestion showed it to be full-length except for 3 or 4 amino acids (Met-Ala-Lys-Arg) at the amino terminus (Fig. 3C). This truncation is ascribed to the trypsin treatment. Importantly, however, there is no indication for cleavage further downstream in or after the hydrophobic region. We conclude that the truncated pro-SI can be translocated into the lumen of the ER and that it retains its hydrophobic signal-anchor sequence.

Construction and Processing of Mutant Signal-Anchor Sequences—No site near the amino terminus of pro-SI shows a particularly high cleavage probability when analyzed with the method of von Heijne (1986a), although three positions have scores of 4.0–4.2. We constructed a series of double mutations leading to sites with high scores, typical of cleaved signals, at various positions in and following the h region.

In the first series of experiments, we considered the hydrophobic stretch as a potentially normal h region and made alterations in the following amino acids expected to improve the following c region. The most favorable cleavage site in

this region lies after position Ala-37 (score of 4.1), but in this case neither Pro-36 nor Lys-34 represent amino acids often found at their positions relative to the cleavage site (von Heijne, 1986a). The double mutant K34G/P36Q (Fig. 1B) has a score of 9.1 for cleavage after position Ala-37; further, the helix-breaking glycine at position -4 relative to the potential cleavage site is expected to be especially favorable (Nothwehr and Gordon, 1989). Although the mutant protein was translocated, the signal sequence was not cleaved (not shown). The single mutation P36Q (Fig. 1C) also gave a translocation-competent but uncleaved protein (not shown).

In the mutations I28P/T33Q and L27A/A29P (Fig. 1, D and E), the scores for cleavage after Ala-32 and Ala-27 were now 7.6 and 7.4, respectively. (As a secondary consideration, in both cases a proline was substituted into the h region, on the chance that it would define the beginning of a c region. See the Introduction). Both of the resulting signal sequences could be more than 90% cleaved (Fig. 2D and not shown). The cleavage sites were determined using radiosequencing with 14 C-labeled proline or alanine. The analysis of mutant I28P/T33Q (Fig. 4, A and B) showed cleavage follows residue Ala-32, the site with the new high score. L27A/A29P was cleaved after Ala-27, again the site with the new high score (Fig. 4, C and D). The two point mutations comprising L27A/A29P (Fig. 1, F and G) were also examined singly. Mutant L27A (with scores 4.8 and 5.4 at positions 26 and 27, respectively) was translocated but not cleaved (not shown). Surprisingly, mutant A29P was cleaved, albeit less efficiently than the double mutant (54 versus $\geq 95\%$). Sequence analysis (Fig. 4, E and F) showed the cleavage occurred after Ala-26 (score of 4.2, same as in the wild type). Thus, the incorporation of a single proline into the h region could convert the uncleaved signal-anchor of sucrase-isomaltase into a cleaved signal, with the cleavage site lying 3 residues upstream of the site of mutation. Note that the cleavage score was not increased by this mutation.

Could the influence of proline on cleavage at upstream positions be more general? We constructed three mutations incorporating proline at position +2 relative to potential signal peptidase cleavage sites: I28P, K34P, and E39P (Fig. 1, H, I, and J). Only I28P was cleaved, and radiosequencing showed the cleavage site to lie between Ala-26 and Leu-27 (Fig. 4, G and H; score for this site is 5.4). Here again, introduction of a single proline led to a cleaved signal sequence. It is noteworthy that in the double mutation I28P/T33Q discussed above, in which Pro-28 is combined with Gln-33, the cleavage site apparently was shifted to a more favorable position further downstream, after Ala-32.

In agreement with the results found with mutant I28P, a deletion mutant, $\Delta 27$ -34 (Fig. 1K), in which Pro-34 of wild type now occupied position 28, also led to cleavage after Ala-26 (score of 4.6) (Fig. 4, I and J). However, the shorter h region in the deletion mutant may also be important. Such an effect is suggested by the observation that another deletion of the h region, $\Delta 23$ -30 (Fig. 1L), showed cleavage mainly after Ala-29 (corresponding to Ala-37 in the wild type sequence), a relatively unfavorable position 1 residue downstream from a proline and having a low cleavage score, 3.2 (Fig. 4, K and L). Some of the product may have been cleaved after Ala-24 (corresponding to Ala-32 in the wild type), which has an even lower cleavage score, 1.5.

Finally, we found that a mutation shortening the n region (mutant $\Delta 4$ -11; Fig. 1M) did not lead to a translocated product (not shown).

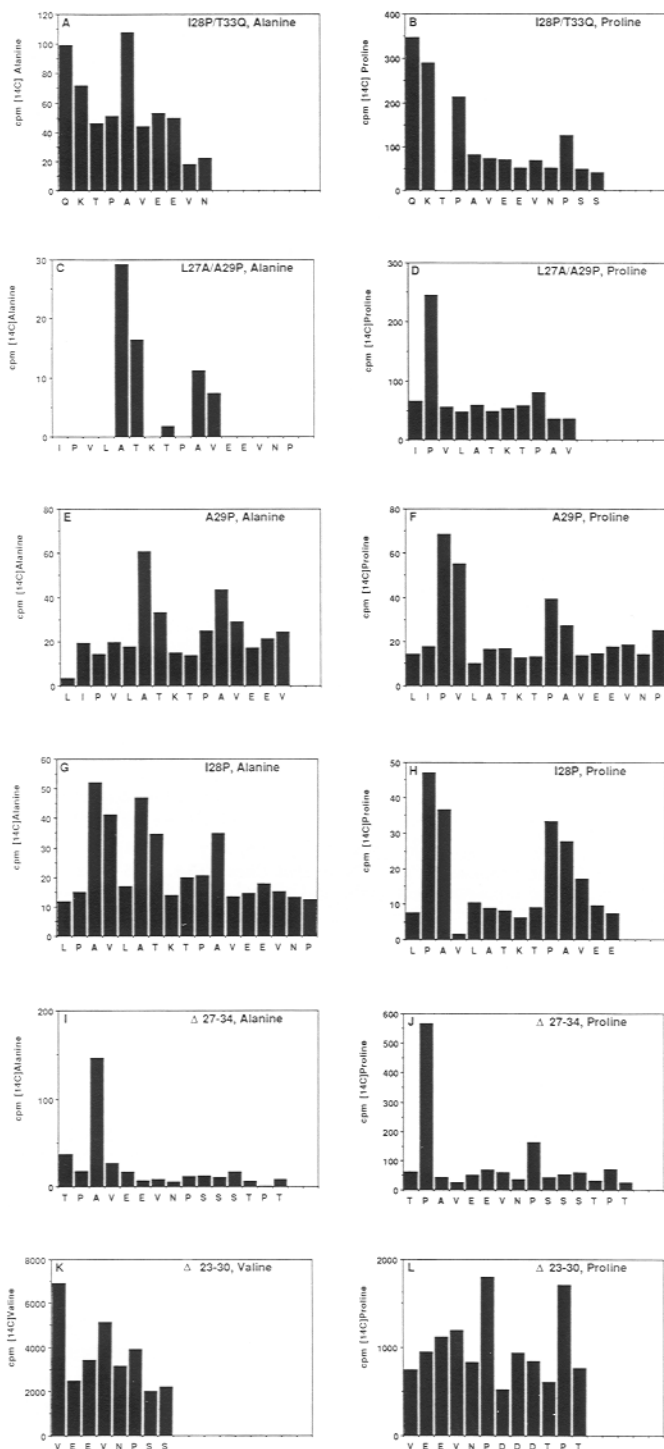


FIG. 4. Radiosequencing of cotranslationally processed mutants of *pro* Δ SI. *In vitro* translation was carried out separately with [14 C]-labeled alanine, proline, or valine in the presence of microsomes, nontranslocated products degraded with trypsin, and the processed polypeptides isolated by gel electrophoresis. After blotting, regions of interest were excised and subjected to automated Edman degradation, and the released radioactivity was counted. The sequences best fitting the data are given on the *abscissa*.

DISCUSSION

The algorithm of von Heijne (1986a) predicts the site of cleavage by signal peptidase in approximately 80% of cleaved signal sequences. A starting point for our work was the realization that the uncleaved signal of *pro*-SI did show at least borderline scores at three positions and that the un-

cleaved signals of the asialoglycoprotein receptor H1 and endopeptidase 24.11 (to pick two of many possible examples) exhibit scores higher than 8. Clearly, other factors are important in determining a cleavage site.

In some cases, introduction of mutations that increase the cleavage score was successful. Improving the score after Ala-32 in the mutant I28P/T33Q, or after Ala-27 in the mutant L27A/A29P, gave cleavable signal sequences, with cleavage occurring at the sites with the new high scores. Further analysis showed, however, that cleavage sites can be activated by single, downstream mutations, in at least one case with no change in cleavage score. Both I28P and, to a lesser extent, A29P were susceptible to signal peptidase, with cleavage occurring *upstream* of the proline substitutions, after Ala-26. In the case of A29P, the score for cleavage after Ala-26 is the same as in the wild type, yet only the mutant signal is cleaved. With mutant L27A, in contrast, the score for this site is raised, but cleavage does not occur. We conclude that introduction of a proline in particular can strongly affect cleavage at more amino-terminal proximal sites.

How can the effect of the proline be explained? The signal-anchor sequence of rat dipeptidylpeptidase IV (Hong and Doyle, 1987) contains a proline, demonstrating that proline *per se* can be tolerated in a signal-anchor sequence. We suspect that a second requirement is that a potential cleavage site be nearby; in mutant I28P, cleavage is virtually complete ($\geq 95\%$) and is observed just 2 amino acids upstream of the introduced proline at a site with a score of 5.4. In mutant A29P, the extent of cleavage drops to about 54%, at a site 3 residues upstream of the introduced proline and with a score of 4.2. In the case of the *uncleaved* signal-anchor of dipeptidylpeptidase IV, the nearest site with a score greater than +1 is after Thr-19 (score of 4.5), 5 amino acids upstream of the proline at position 24.

An effect of a downstream proline has also been described for signal peptidase I of *E. coli*. Duffaud and Inouye (1988) analyzed a series of mutants of an OmpA-nuclease hybrid in which a Pro at +3 from the cleavage site was deleted. In some mutants this was combined with other changes predicted to further decrease the probability of a turn structure. Progressively reduced processing was observed as more changes were introduced, suggesting that signal peptidase recognizes a feature of the secondary structure such as a turn at the cleavage site (Inouye *et al.*, 1986; Duffaud and Inouye, 1988). Our own results are consistent with the idea that cleavage by mammalian signal peptidase is also favored by such a feature.

Also, the length of the h region clearly exerts an important influence on signal peptidase. The double mutant K34G/P36G produces a score >9 at Ala-37, some 5 amino acids downstream of the 20-amino acid h region. However, it is not cleaved. This is analogous to the situation in the asialoglycoprotein receptor. By contrast, shortening the h region, as in *pro*-SI Δ 23-30, leads to cleavage after Ala-29 (corresponding to Ala-37 in the wild type), even though the score at this position is only 3.2. In this mutant, the h region is reduced to only 12 amino acids, more typical of a cleaved signal (von Heijne, 1985; Briggs and Gierasch, 1986). Our mutant *pro*-SI Δ 27-34 has an h region 14 amino acids long and is also cleaved, after Ala-26 (score of 4.6). However, based on the results with I28P, the presence of a proline at +2 from the cleavage site in *pro*-SI Δ 27-34 is also likely to play a role in this mutant. It is noteworthy that shortening the h region of dipeptidylpeptidase IV to 15 amino acids did *not* lead to a cleaved signal, even though a score of 4.6 was found at the end of the new h region (mutant "M4" in Hong and Doyle (1990)).

von Heijne (1985) has analyzed the hydrophobicity of a set of cleaved signal sequences as a function of the distance from the cleavage site and concluded that there is on the whole a clear difference between the h region and the c region, the c region being more polar. However, in three of our cleaved mutants, I28P, A29P, and $\Delta 27-34$, what has by definition become the c region, in terms of being the 5 or so amino acids upstream from the new cleavage site, is in fact unchanged in sequence from the wild type, where the segment formed part of the h region. Thus, there are no special sequence characteristics, either hydrophobicity or otherwise, which necessarily differentiate a c region from part of an h region.

Drastic shortening of the n region, which led to cleavage of the immunoglobulin invariant chain (Lipp and Dobberstein, 1986) and of the asialoglycoprotein receptor H1 (Schmid and Spiess, 1988), in our system did not lead to a translocated product. This might be ascribed to a blockage in translocation, as has been observed in deletions of the n region of a parainfluenza virus hemagglutinin-neuraminidase (Spriggs and Collins, 1990). Alternatively, the polarity of translocation may have been reversed by the change in the balance of charged residues upstream and downstream of the n region (Hartmann *et al.*, 1989), as has been seen with, for example, derivatives of the asialoglycoprotein receptor (Beltzer *et al.*, 1991) and paramyxovirus hemagglutinin-neuraminidase (Parks and Lamb, 1991).

In conclusion, we have found that surprisingly small changes can convert the signal-anchor of pro-SI to a cleaved signal sequence. A proline *downstream* of a potential cleavage site appears particularly favorable, perhaps by influencing the secondary structure of the polypeptide in this region. The length of the h region also plays a major role; cleavage can occur even at a relatively unfavorable site if the preceding h region is shortened, whereas simply improving the cleavage score of the site is insufficient. The exact nature of the c region is not necessarily important, since the same sequence can function as a c region in one context and as part of an h region in another.

Acknowledgments—We thank Jürg Ruf for advice and help with the translation/translocation system and Werner Boll and Martin Spiess for synthesis of oligonucleotides.

REFERENCES

- Anderson, C. W., Straus, J. W., and Dudock, B. S. (1983) *Methods Enzymol.* **101**, 635–644.
- Andrews, D. W., Perara, E., Lesser, C., and Lingappa, V. R. (1988) *J. Biol. Chem.* **263**, 15791–15798.
- Beltzer, J. P., Wessels, H. P., and Spiess, M. (1989) *FEBS Lett.* **253**, 93–98.
- Beltzer, J. P., Fiedler, K., Fuhrer, C., Geffen, L., Handschin, C., Wessels, H. P., and Spiess, M. (1991) *J. Biol. Chem.* **266**, 973–978.
- Bos, T. J., Davis, A. R., and Nayak, D. P. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 2327–2331.
- Briggs, M. S., and Gierasch, L. M. (1986) *Adv. Protein Chem.* **38**, 109–180.
- Burgess, T. L., Craik, C. S., Matsuuchi, L., and Kelly, R. B. (1987) *J. Cell Biol.* **105**, 659–668.
- Cioffi, J. A., Allen, K. L., Lively, M. O., and Kemper, B. (1989) *J. Biol. Chem.* **264**, 15052–15058.
- Duffaud, G., and Inouye, M. (1988) *J. Biol. Chem.* **263**, 10224–10228.
- Folz, R. J., and Gordon, J. I. (1986) *J. Biol. Chem.* **261**, 14752–14759.
- Folz, R. J., and Gordon, J. I. (1987) *J. Biol. Chem.* **262**, 17221–17230.
- Folz, R. J., Nothwehr, S. F., and Gordon, J. I. (1988) *J. Biol. Chem.* **263**, 2070–2078.
- Ghera, P., Huber, P., Semenza, G., and Wacker, H. (1986) *J. Biol. Chem.* **261**, 7969–7974.
- Hartmann, E., Rapoport, T. A., and Lodish, H. F. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 5786–5790.
- Hong, W., and Doyle, D. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 7962–7966.
- Hong, W., and Doyle, D. (1990) *J. Cell Biol.* **111**, 323–328.
- Hunziker, W., Spiess, M., Semenza, G., and Lodish, H. F. (1986) *Cell* **46**, 227–234.
- Inouye, S., Duffaud, G., and Inouye, M. (1986) *J. Biol. Chem.* **261**, 10970–10975.
- Jagus, R. (1987) *Methods Enzymol.* **152**, 267–276.
- Kramer, W., and Fritz, H. J. (1987) *Methods Enzymol.* **154**, 360–367.
- Kramer, W., Drutsa, V., Jansen, H. W., Kramer, B., Pflugfelder, M., and Fritz, H. J. (1984) *Nucleic Acids Res.* **12**, 9441–9456.
- Kramer, W., Ohmayer, A., and Fritz, H. J. (1988) *Nucleic Acids Res.* **16**, 7207.
- Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 488–492.
- Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382.
- Laemmli, U. K. (1970) *Nature* **227**, 680–685.
- Lipp, J., and Dobberstein, B. (1986) *Cell* **46**, 1103–1112.
- Melton, D. (1987) *Methods Enzymol.* **152**, 288–295.
- Mumford, R. A., Pickett, C. B., Zimmerman, M., and Strauss, A. W. (1981) *Biochem. Biophys. Res. Commun.* **103**, 565–572.
- Nelson, R. M., and Long, G. L. (1989) *Anal. Biochem.* **180**, 147–151.
- Nothwehr, S. F., and Gordon, J. I. (1989) *J. Biol. Chem.* **264**, 3979–3987.
- Nothwehr, S. F., Hoeltzli, S. D., Allen, K. L., Lively, M. O., and Gordon, J. I. (1990) *J. Biol. Chem.* **265**, 21797–21803.
- Parks, G. D., and Lamb, R. A. (1991) *Cell* **64**, 777–787.
- Payvar, F., and Schimke, R. T. (1979) *J. Biol. Chem.* **254**, 7636–7642.
- Perara, E., and Lingappa, V. R. (1985) *J. Cell Biol.* **101**, 2292–2301.
- Schägger, H., and von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379.
- Schmid, S. R., and Spiess, M. (1988) *J. Biol. Chem.* **263**, 16886–16891.
- Siegel, V., and Walter, P. (1985) *J. Cell Biol.* **100**, 1913–1921.
- Spriggs, M. K., and Collins, P. (1990) *J. Cell Biol.* **111**, 31–44.
- Vlasuk, G. P., Inouye, S., and Inouye, M. (1984) *J. Biol. Chem.* **259**, 6195–6200.
- von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21.
- von Heijne, G. (1984a) *EMBO J.* **3**, 2315–2318.
- von Heijne, G. (1984b) *J. Mol. Biol.* **173**, 243–251.
- von Heijne, G. (1985) *J. Mol. Biol.* **184**, 99–105.
- von Heijne, G. (1986a) *Nucleic Acids Res.* **14**, 4683–4690.
- von Heijne, G. (1986b) *J. Mol. Biol.* **189**, 239–242.
- Walter, P., and Blobel, G. (1983) *Methods Enzymol.* **96**, 84–111.
- Walter, P., Ibrahim, I., and Blobel, G. (1981) *J. Cell Biol.* **91**, 545–561.
- Waters, M. G., Evans, E. A., and Blobel, G. (1988) *J. Biol. Chem.* **263**, 6209–6214.
- Wessel, D., and Flügge, U. I. (1984) *Anal. Biochem.* **138**, 141–143.
- Yamamoto, Y., Taniyama, Y., and Kikuchi, M. (1989) *Biochemistry* **28**, 2728–2732.