THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 267, No. 24, Issue of August 25, pp. 16928–16933, 1992
Printed in U.S.A.

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. 1*A*. 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.

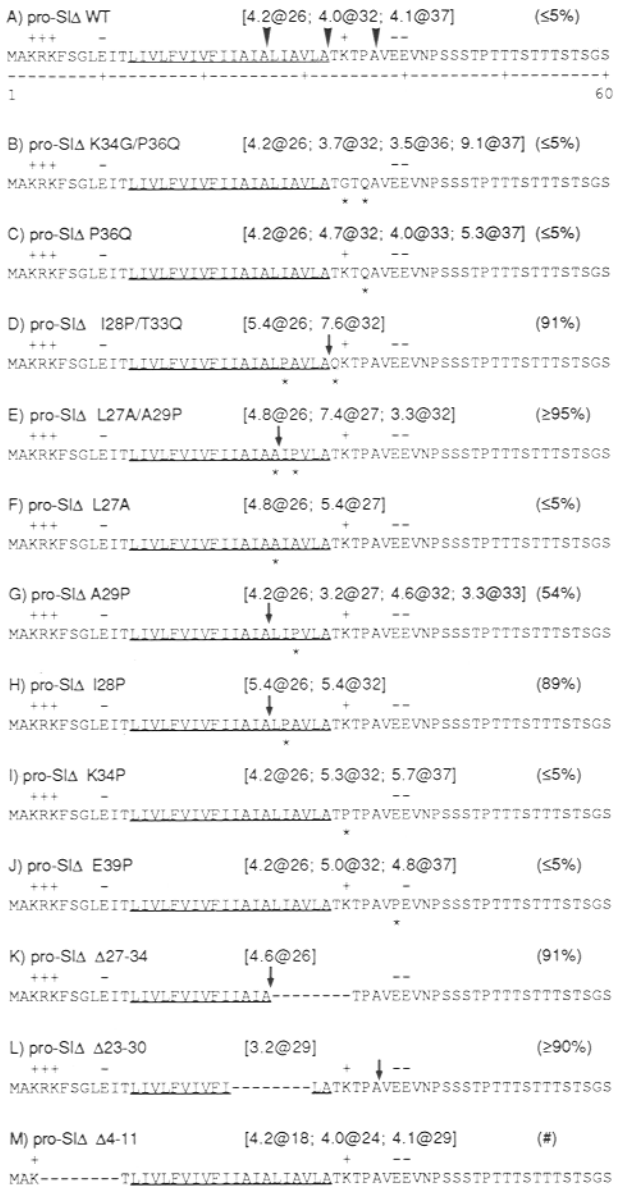

FIG. 1. **Sequence of the N-terminal proximal region of rabbit sucrase-isomaltase coded by pro-SIΔ wild type and derivatives.** Charged amino acids are indicated by *plus* and *minus signs*; the hydrophobic transmembrane region is *underlined*; amino acids changed in mutants are indicated with *asterisks*; 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(Δpro)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 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-^{14}C]Alanine (>150 mCi/mmol) and L-[U-^{14}C]proline (>290 mCi/mmol) were from Amersham Corp. L-[^{35}S]Cysteine (1100 mCi/mmol), L-[3,4,5-^{3}H]leucine (156 mCi/mmol), and L-[U-^{14}C]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-SIΔ carries a segment of wild-type pro-SI cDNA (Hunziker *et al.*, 1986) that had been originally subcloned into pGem-4, from a *Bam*HI site (18 base pairs upstream of the initiation codon) to a *Bgl*II site 701 base pairs downstream. A sequence derived from the original pGem-4 polylinker region another 8 base pairs downstream from the *Bgl*II site introduces an artificial stop codon just before a *Sal*I 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 *Sal*I was transcribed with T7 RNA polymerase in the presence of 0.5 mM m^{7}pGpppG (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.