Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 16–19, January 1992
Biochemistry

Functions of signal and signal–anchor sequences are determined by the balance between the hydrophobic segment and the N-terminal charge

(protein translocation/membrane protein/cytochrome P-450)

MASAO SAKAGUCHI*, RYOKO TOMIYOSHI†, TOMOKO KUROIWA†, KATSUYOSHI MIHARA*, AND TSUNEO OMURA*‡

*Department of Molecular Biology, Graduate School of Medical Science and †Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812 Japan

Communicated by David D. Sabatini, October 2, 1991 (received for review May 28, 1991)

ABSTRACT **The signal sequence of secretory proteins and the signal–anchor sequence of type II membrane proteins initiate the translocation of the following polypeptide segments, whereas the signal–anchor sequence of cytochrome P-450-type membrane proteins mediates the membrane insertion of the polypeptide via a signal-recognition particle-dependent mechanism but does not lead to the translocation of the following C-terminal sequences. To establish the structural requirements for the function of signal and signal–anchor sequences, we constructed chimeric proteins containing artificial topogenic sequences in which the N-terminal net charge and the length of the hydrophobic segment were systematically altered. Utilizing an *in vitro* translation–translocation system, we found that hydrophobic segments consisting of 7–10 leucine residues functioned as signal sequences whereas segments with 12–15 leucine residues showed different topogenic functions, behaving as signal sequences or P-450-type signal–anchor sequences, depending on the N-terminal charge. From these observations, we propose that the function of N-terminal topogenic sequences depends on a balance between the N-terminal charge and the length of the following hydrophobic segment.**

Many integral membrane proteins of the endoplasmic reticulum (ER)-derived cell organelles are synthesized on membrane-bound ribosomes and are cotranslationally inserted into the ER membrane (1, 2). Their topologies in the membrane are determined by so-called topogenic sequences (1, 3) in their primary structures. The topogenic sequences are categorized as follows. (*i*) Signal sequences (S) are located at the N termini of the nascent polypeptides of secretory proteins and type I membrane proteins and mediate initiation of peptide translocation across the membrane. These sequences are cleaved by the signal peptidase on the luminal side of ER membrane. (*ii*) Stop–transfer sequences interrupt the translocation and anchor the protein in the membrane. The sequential action of an S sequence and this type of topogenic sequence establishes the N_{exo}/C_{cyt} orientation of type I membrane proteins. (*iii*) Signal–anchor sequences termed SAII are found in type II membrane proteins, which possess a single transmembrane segment with an N_{cyt}/C_{exo} orientation. These are uncleavable signal sequences that mediate the translocation of the following portions of the polypeptide and anchor the protein in the membrane. SAII sequences are categorized as variants of the classical S sequence. Their functional difference from S is thought to result from the absence of an accessible site for the signal peptidase (4, 5). Transferrin receptor (6), invariant chain of class II histocompatibility antigen (4), asialoglycoprotein receptor (7), and hemagglutinin/neuraminidase (8) are examples of type II membrane proteins. (*iv*) Sequences tentatively named type I signal–anchor sequences (SAI) are present in P-450-type membrane proteins and are short N-terminal segments that have the combined topogenic functions of a signal sequence and a stop–transfer sequence (9–11). They lead to a type I orientation (N_{exo}/C_{cyt}) of the protein in the membrane (10). Microsomal cytochrome P-450 proteins (12), NADPH cytochrome P-450 reductase (13), and the M_2 protein of the influenza A virus (14) possess SAI sequences. P-450 molecules are cotranslationally inserted into the membrane by a process mediated by the signal-recognition particle (15, 16) and are anchored only by the N-terminal signal–anchor sequence (12, 17); the remainder of the protein is exposed on the cytoplasmic surface of the membrane (9, 12, 18). The final disposition of SAI is inverted in its orientation compared with that of the other two sequences, S and SAII (Fig. 1).

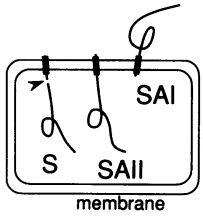

FIG. 1. Classification of membrane topology of signal and signal–anchor sequences in ER membrane. S indicates the cleavable signal sequence. SAI and SAII represent signal–anchor sequences of P-450-type membrane proteins and type II membrane proteins, respectively. Arrowhead indicates the processing by signal peptidase.

Both SAI and SAII sequences have been demonstrated to be targeted to the ER membrane by the signal-recognition particle, as is the S sequence (9, 14, 19). One can therefore ask what structural features determine the orientation of these topogenic sequences in the membrane.

SAI of P-450 can be converted to SAII or S by replacing the N-terminal negatively charged amino acid residues preceding the hydrophobic core of the signal with positively charged ones (11, 20, 21). This is in accord with the proposal that the charged residues flanking the hydrophobic segment of a topogenic sequence are an important determinant of the orientation of the hydrophobic stretch in the membrane, with the flanking side that is rich in positively charged residues located on the cytoplasmic side (22–26). This finding also establishes the functional importance of the negatively charged residues that almost all microsomal P-450s possess in

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ER, endoplasmic reticulum; RM, rough microsomal.

‡To whom reprint requests should be addressed.

their N-terminal portions (12, 27). From this perspective, the orientation of the hydrophobic segment would be defined by the relative ability of the two flanking portions to be translocated. Recently, the introduction of positively charged residues placed downstream of a hydrophobic segment was shown to enhance the capacity of the segment to halt translocation of downstream portions of the polypeptide (28, 29).

However, the SAI sequence of P-450 was also converted to S by partial deletion of the hydrophobic segment, even when the N-terminal negatively charged residues were retained (21). This suggested that another factor(s) besides the flanking charged residues participates in the determination of the orientation of topogenic sequences. We have evaluated the contributions of the length of the hydrophobic segment and the N-terminal charged residues to the topogenic function by using systematically constructed model sequences. Long hydrophobic segments functioned as S when positive charges were present at the N terminus and functioned as SAI when the charge was negative. Shorter hydrophobic segments functioned as S even when the N-terminal charge was negative. We propose that the functions of the S and SAI sequences depend on the balance between the length of the hydrophobic segment and the N-terminal charge.

MATERIALS AND METHODS

Materials. Plasmid pSPIL22, in which cDNA of mouse interleukin 2 has been inserted downstream of the SP6 RNA polymerase promoter in pGEM-2 vector (Promega), has been described (28, 29). Rough microsomal (RM) membranes of dog pancreas (30) and wheat germ cell-free extracts (S-23 fraction) (31) were prepared as described. SP6 RNA polymerase, RNase inhibitor, restriction enzymes, and DNA-modifying enzymes were purchased from Toyobo (Osaka) and Takara Shuzo (Kyoto). Proteinase K was obtained from Merck. ^{35}S protein labeling mix ([^{35}S]methionine, 43.0 TBq/mmol) was purchased from NEN.

Construction of Plasmids. pSPIL22 was digested with *Hin*cII and self-ligated to delete the sequence encoding N-terminal 18 amino acid residues from the signal peptide, which consists of 20 amino acid residues. The *Hin*dIII site in the coding region of interleukin 2 was destroyed by partial *Hin*dIII digestion, filling-in, and subsequent ligation of an *Xho* I linker (octanucleotide) to generate pDIL2. The following three sets of oligonucleotides, encoding the initiation methionine and charged residues and including *Hin*dIII and *Bam*HI restriction enzyme sites, were inserted between the *Eco*RI and *Xba* I sites of pDIL2: (*i*) 5′-d(AATTCATGGATCCAAGCTTAT)-3′ and 3′-d(GTACCTAGGTTCGAATAGATC)-5′; (*ii*) 5′-d(AATTCATGCGGCGGGATCCAAGCTTAT)-3′ and 3′-d(GTACGCCGCCCTAGGTTCGAATAGATC)-5′; (*iii*) 5′-d(AATTCATGCGCCGGCGGCGGGATCCAAGCTTAT)-3′ and 3′-d(GTACGCGGCCGCCGCCCTAGGTTCGAATAGATC)-5′ (the sequences coding for N-terminal amino acid residues are underlined). Oligonucleotides encoding leucine clusters were synthesized and inserted between *Bam*HI and *Hin*dIII sites in the above oligonucleotides, as described (28). All recombinants were screened by restriction enzyme mapping and confirmed by sequencing. The N-terminal amino acid sequences of the constructed proteins are shown in Fig. 2, and the proteins are designated according to the inserted hydrophobic segment and N-terminal net charge. For example, the L15(−) molecule has 15 leucines in the hydrophobic segment and a net charge of −1 in the N-terminal region.

***In Vitro* Assay of Topogenic Function of Model Sequences.** The constructed cDNAs were transcribed by SP6 RNA polymerase, essentially as described (32). DNA templates were prepared from overnight cultured cells by SDS/alkali extraction and subsequent RNase treatment (33). Plasmid DNAs were precipitated with polyethylene glycol and digested with *Pvu* I.

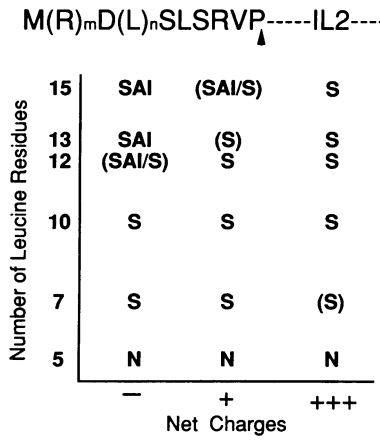

FIG. 2. Summary of constructed N-terminal sequences and their topogenic functions. Number of N-terminal arginine residues (indicated by m) was 0, 2, or 4. Number of leucine residues (indicated by n) was 5, 7, 10, 12, 13, or 15. IL2 represents the interleukin 2 sequence. S, SAI, and N indicate signal sequence, signal–anchor sequence of P-450-type membrane protein, and no interaction with the membrane, respectively. Symbols in parentheses indicate weak function of the topogenic sequences.

Messenger RNAs transcribed from *Pvu* I-digested plasmids were translated at 26°C for 45 min in a wheat germ cell-free system in the presence or absence of RM membranes (1.5 equivalents per 25 μl). Aliquots (10 μl) were treated with proteinase K (200 μg/ml) at 20°C for 30 min. The digestion was terminated by adding 200 μl of 10% (wt/vol) trichloroacetic acid and proteins were recovered by centrifugation.

Other aliquots were mixed with 100 μl of 0.1 M Na_2CO_3 (pH 11.4) and incubated at 0°C for 15 min before centrifugation in a Beckman Airfuge for 90 sec at 30 psi (206.7 kPa) to precipitate the membrane (9). The supernatant was mixed with 100 μl of 20% trichloroacetic acid to precipitate the extracted proteins. Proteins were analyzed by SDS/15% polyacrylamide gel electrophoresis (34) and subsequent fluorography (35).

Other Methods. DNA manipulations were performed according to standard protocols (33). The nucleotide sequences of manipulated regions were confirmed by dideoxy chain termination with alkali-denatured double-stranded plasmid DNAs as templates (33, 36). Oligonucleotides were synthesized with an automated DNA synthesizer (model 380B, Applied Biosystems).

RESULTS

To precisely assess the contribution of the charge and the length of the hydrophobic segment to the function of topogenic sequences, we constructed systematically engineered model topogenic sequences and linked them to the interleukin 2 polypeptide, thus replacing the natural signal for this secretory protein (9). In the series of model sequences, the net N-terminal charge (−1, +1, and +3) and the length of the following hydrophobic segment (5, 7, 10, 12, 13, and 15 leucine residues) were systematically altered (Fig. 2). The original processing site for the signal peptidase (37) was retained in the chimeric proteins.

The constructed cDNAs were transcribed *in vitro* with SP6 RNA polymerase and translated in the presence or absence of RM membranes (Fig. 3). Precursor polypeptides with uncleaved signals were detected as single radioactive bands when the translation was carried out in the absence of RM

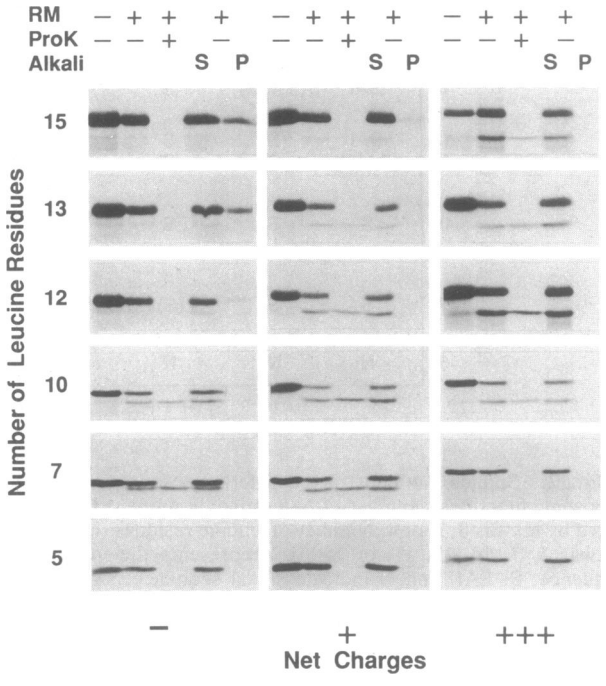

FIG. 3. Topogenic functions of the constructed sequences. These results are summarized in Fig. 2. Messenger RNAs transcribed *in vitro* were translated with a wheat germ cell-free system in the presence (+) or absence (−) of RM membranes (1.5 equivalents per 25 μl). An aliquot (10 μl) was treated with proteinase K (ProK, 200 μg/ml) at 20°C for 30 min. Another aliquot was treated with 0.1 M Na_2CO_3 (pH 11.4; alkali) at 0°C for 15 min and the supernatant (S) and membrane pellet (P) fractions were separated. Proteins were analyzed by SDS/polyacrylamide gel electrophoresis and subsequent fluorography.

membranes (RM − lanes). When translation was carried out in the presence of RM membranes (RM + lanes), mature polypeptides resulting from signal cleavage were observed with some constructs. Upon posttranslational treatment with proteinase K (ProK + lanes), these mature (processed) polypeptides were not degraded whereas the precursor forms were completely degraded. The mature polypeptides were, however, degraded by the proteinase when the membranes were solubilized with detergent (data not shown). These data confirmed that the mature polypeptides were translocated across the RM membranes and, moreover, that the model sequences displayed an S function. In no case were the precursor bands resistant to the proteinase, indicating that all the peptides that had been translocated were completely processed by the signal peptidase and, therefore, that none of the model sequences showed the topogenic function of SAII.

To examine the extent of membrane anchoring of the various products, the translation mixtures containing RM membranes were treated with 0.1 M Na_2CO_3 and the membrane pellet and supernatant fractions were separated (lanes P and S in Fig. 3). Under these conditions, all peripheral membrane proteins and soluble luminal proteins are extracted (9, 38), so that both the precursor and mature forms of the polypeptides of secretory proteins are recovered in the supernatant. On the other hand, when the model sequence functioned as SAI, the precursor polypeptides would remain anchored to the membrane and substantial portions would be recovered in the membrane fraction (P). Such polypeptides, however, would be sensitive to the proteinase, even in the absence of detergent. The treatment with Na_2CO_3 seems to be a reliable criterion for membrane integration of the polypeptides. Indeed, in our previous studies (28, 29) on the function of stop–transfer sequences the recoveries of the translocation-interrupted polypeptides in the membrane fraction after alkali extraction were about 50% and were independent of the lengths of the hydrophobic segments (which varied from 6 leucine to 19 alanine residues).

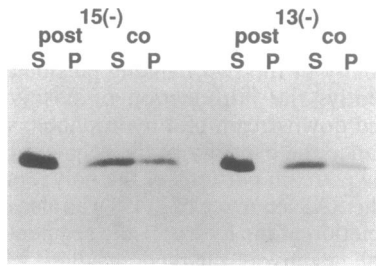

FIG. 4. Cotranslational binding of L15(−) and L13(−) molecules to RM membranes. The mRNAs for L15(−) and L13(−) proteins were translated with wheat germ cell-free system in the absence (post) or in the presence (co) of RM membranes (1.5 equivalents per 25 μl). After the termination of the translation with cycloheximide (1 mM), RM membranes were added to the translation mixture without RM membranes (post) and the mixtures were further incubated at 26°C for 45 min. The translation mixtures were treated with the "alkali" (Na_2CO_3) solution, and the supernatant and the membrane precipitates were separated and analyzed as described in the legend to Fig. 3.

The topogenic functions of the 18 constructed model sequences depended both on the charge and on the length of the hydrophobic segment (Figs. 2 and 3). L15(−) and L13(−) molecules were not translocated and substantial portions of the precursor polypeptides were recovered in the membrane fraction (Fig. 3, P lanes), indicating that the sequences functioned as SAI. When RM membranes were added after the completion of the translation reaction, L15(−) and L13(−) molecules were not recovered in the membrane fractions (Fig. 4, "post" lanes). This result demonstrated that the membrane binding of these molecules was achieved only cotranslationally and was not due to nonspecific adsorption to the membranes. In the case of L12(−) and L15(+) molecules, both proteinase K-resistant processed polypeptides (Fig. 3, ProK + lanes) and alkali-resistant mature-size polypeptides (Fig. 3, P lanes) were observed, indicating that the functions of the sequences were on the threshold between S and SAI.

Model sequences containing more positively charged residues and/or shorter hydrophobic segments functioned as S. The efficiency of the processing varied depending on the net charge of the sequence and the length of the leucine cluster. Whenever processing occurred, cleavage appeared to occur at the authentic processing site of interleukin 2, since the mature polypeptides all showed the same electrophoretic mobility. The model sequences containing the shortest hydrophobic segments no longer interacted with the membrane. This is in accord with previous reports that the hydrophobicity of the core segment in the S sequence is directly related to the topogenic function of the signal (39, 40).

DISCUSSION

The model sequences analyzed in this study showed different topogenic functions, S or SAI. Those containing 7–10 leucine residues functioned as S irrespective of their N-terminal charge. When the hydrophobic segment was longer, however, the function of the sequences was dependent on the N-terminal charge. Our data clearly demonstrate that the function of topogenic sequences can be determined by a balance between the net charge preceding the hydrophobic segment and the length of the segment.

N-terminal positively charged residues enabled the topogenic sequences containing longer hydrophobic segments to function as S, and the number of positive charges required for

S function increased as the hydrophobic stretch was elongated. This seems to explain the fact that positive charges are found at the N terminus in the majority of eukaryotic S-type signal sequences (41). Although the positive charges of such signal sequences are not as obligatory in the eukaryotic ER membrane system as in *Escherichia coli* protein-export system (9, 40, 42), they obviously improve the function of S sequences. We propose that the positive charge acts by fixing the N-terminal portion of the topogenic sequences on the cytoplasmic surface of the membrane and, consequently, enhances the translocation of the C-terminal portion across the membrane. A similar proposal was made by Parks and Lamb (25), who studied the effect of N-terminal charges on the orientation of influenza virus envelope proteins. The role of the positive charges as an energy barrier to the translocation of a polypeptide was also apparent from our studies of the structural requirements for the stop–transfer activity of a topogenic sequence (28, 29).

It was demonstrated in this study that hydrophobic segments as short as 7–10 leucine residues (core hydrophobic segment, or h region; ref. 41) were sufficient for the function of S and that this function was not affected by the N-terminal charge. In contrast, the function of longer hydrophobic segments was dependent on the preceding charge. It is tempting to speculate that the capacity of a polypeptide segment to function as a eukaryotic S-type signal sequence is determined primarily by the core hydrophobic segment, which can loop into the protein-translocation machinery in the membrane, and that when the hydrophobic segment is longer it will tend to pull the N-terminal portion into the membrane and the orientation of SAI will be achieved as in the case of L15(−) and L13(−) molecules. Positively charged residues of the N-terminal portion would provide a force preventing the N terminus from entering the membrane (25, 28, 29). The balance between the pulling force and the resisting force determines the ultimate location of the N terminus.

We thank Sueyoshiseifun, Nara, Japan, for providing us fresh wheat germ. This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan.

1. Sabatini, D. D., Kreibich, G., Morimoto, T. & Adesnik, M. (1982) *J. Cell Biol.* **92,** 1–22.
2. Wickner, W. T. & Lodish, H. F. (1985) *Science* **230,** 400–407.
3. Blobel, G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1496–1500.
4. Lipp, J. & Dobberstein, B. (1986) *Cell* **46,** 1103–1112.
5. Shaw, A. S., Rottier, P. J. M. & Rose, J. K. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7592–7596.
6. Zerial, M., Melancon, P., Schneider, C. & Garoff, H. (1986) *EMBO J.* **5,** 1543–1550.
7. Spiess, M. & Lodish, H. F. (1986) *Cell* **44,** 177–185.
8. Hiebert, S. W., Paterson, R. G. & Lamb, R. A. (1985) *J. Virol.* **54,** 1–6.
9. Sakagachi, M., Mihara, K. & Sato, R. (1987) *EMBO J.* **6,** 2425–2431.
10. Monier, S., Luc, P. V., Kreibich, G., Sabatini, D. D. & Adesnik, M. (1988) *J. Cell Biol.* **107,** 457–470.
11. Szczesna-Skorupa, E., Browne, N., Mead, D. & Kemper, B. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 738–742.
12. Nelson, D. R. & Strobel, H. W. (1988) *J. Biol. Chem.* **263,** 6038–6050.
13. Black, S. D. & Coon, M. J. (1982) *J. Biol. Chem.* **257,** 5929–5938.
14. Hull, J. D., Gilmore, R. & Lamb, R. A. (1988) *J. Cell Biol.* **106,** 1489–1498.
15. Bar-Nun, S., Kreibich, G., Adesnik, M., Alterman, L., Negishi, M. & Sabatini, D. D. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 965–969.
16. Sakaguchi, M., Mihara, K. & Sato, R. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3361–3364.
17. Brown, C. A. & Black, S. D. (1989) *J. Biol. Chem.* **264,** 4442–4449.
18. Lemos-Chiarandini, C. D., Frey, A. B., Sabatini, D. D. & Kreibich, G. (1987) *J. Cell Biol.* **104,** 209–219.
19. Lipp, J. & Dobberstein, B. (1986) *J. Cell Biol.* **102,** 2169–2175.
20. Szczesna-Skorupa, E. & Kemper, B. (1989) *J. Cell Biol.* **108,** 1237–1243.
21. Sato, T., Sakaguchi, M., Mihara, K. & Omura, T. (1990) *EMBO J.* **9,** 2391–2397.
22. Hartmann, E., Rapoport, T. A. & Lodish, H. F. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5786–5790.
23. Haeuptle, M.-T., Flint, N., Gough, N. M. & Dobberstein, B. (1989) *J. Cell Biol.* **108,** 1227–1236.
24. von Heijne, G. (1989) *Nature (London)* **341,** 456–458.
25. Parks, G. D. & Lamb, R. A. (1991) *Cell* **64,** 777–787.
26. Beltzer, J. P., Fiedler, K., Fuhrer, C., Geffen, I., Handschin, C., Wessels, H. P. & Spiess, M. (1991) *J. Biol. Chem.* **266,** 973–978.
27. Gotoh, O. & Fujii-Kuriyama, Y. (1989) in *Frontiers in Biotransformation*, eds. Ruckpaul, K. & Rein, H. (Akademie, Berlin), pp. 195–245.
28. Kuroiwa, T., Sakaguchi, M., Mihara, K. & Omura, T. (1990) *J. Biochem.* **108,** 829–834.
29. Kuroiwa, T., Sakaguchi, M., Mihara, K. & Omura, T. (1991) *J. Biol. Chem.* **266,** 9251–9255.
30. Walter, P. & Blobel, G. (1983) *Methods Enzymol.* **96,** 84–93.
31. Erickson, A. H. & Blobel, G. (1983) *Methods Enzymol.* **96,** 38–50.
32. Krieg, P. A. & Melton, D. A. (1984) *Nucleic Acids Res.* **12,** 7057–7070.
33. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
34. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
35. Bonner, W. M. & Laskey, R. A. (1974) *Eur. J. Biochem.* **46,** 83–88.
36. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
37. Kashima, N., Nishi-Takaoka, C., Fujita, T., Taki, S., Yamada, G., Hamuro, J. & Taniguchi, T. (1985) *Nature (London)* **313,** 402–404.
38. Fujiki, Y., Hubbard, A. L., Fowler, S. & Lazarow, P. B. (1982) *J. Cell Biol.* **93,** 97–102.
39. Bird, P., Gething, M.-J. & Sambrook, J. (1990) *J. Biol. Chem.* **265,** 8420–8425.
40. Nothwehr, S. F. & Gordon, J. I. (1990) *J. Biol. Chem.* **265,** 17202–17208.
41. von Heijne, G. (1985) *J. Mol. Biol.* **184,** 99–105.
42. Nothwehr, S. F. & Gordon, J. I. (1990) *BioEssays* **12,** 479–484.