

Analysis of endoplasmic reticulum trafficking signals by combinatorial screening in mammalian cells

Noa Zerangue, Michael J. Malan, Sharon R. Fried, Paul F. Dazin, Yuh Nung Jan, Lily Yeh Jan, and Blanche Schwappach*

Howard Hughes Medical Institute, Departments of Physiology and Biochemistry, Program in Neuroscience, University of California, San Francisco, CA 94143-0725

Contributed by Lily Yeh Jan, December 28, 2000

To improve the accuracy of predicting membrane protein sorting signals, we developed a general methodology for defining trafficking signal consensus sequences in the environment of the living cell. Our approach uses retroviral gene transfer to create combinatorial expression libraries of trafficking signal variants in mammalian cells, flow cytometry to sort cells based on trafficking phenotype, and quantitative trafficking assays to measure the efficacy of individual signals. Using this strategy to analyze arginine- and lysine-based endoplasmic reticulum localization signals, we demonstrate that small changes in the local sequence context dramatically alter signal strength, generating a broad spectrum of trafficking phenotypes. Finally, using sequences from our screen, we found that the potency of di-lysine, but not di-arginine, mediated endoplasmic reticulum localization was correlated with the strength of interaction with α -COP.

The compartments of the secretory pathway are linked by a continuous flux of membranes and proteins, yet each compartment maintains a distinct protein composition. To achieve this, individual proteins must be packaged into or excluded from transport vesicles by sorting signals. The best characterized sorting signals include the cytosolic di-lysine (-KKXX-COOH) and luminal -KDEL-COOH signals for endoplasmic reticulum (ER) retention/retrieval (1), and tyrosine-based and di-leucine motifs involved in endocytosis and trans-Golgi network (TGN) trafficking (2, 3). Arginine-based ER retention/retrieval signals are less well characterized (1, 4, 5). Each of these signals requires several key residues, but less is known about how surrounding residues affect their function. The consensus for most trafficking signals has been derived from limited site-directed mutagenesis of a few example proteins. The interaction between the YXX Φ sequence and the adaptor complex μ chain can be reconstituted in the yeast two-hybrid system or *in vitro* binding, allowing combinatorial screens for the sequence consensus (6, 7). However, such an approach provides information about which sequences are optimal for *in vitro* biochemical interaction, but the consensus for *in vivo* trafficking has not been systematically investigated in the same way for any trafficking signal.

To address this problem, we have devised a new method to explore the consensus and context dependence of intracellular sorting signals in living mammalian cells. We find that local sequence context creates a broad spectrum of signals with varying potency. In addition, we show that the strongest -KKXX-signals from our screen bind the α -subunit of the COPI vesicle coat in the yeast two-hybrid assay.

Materials and Methods

Molecular Biology. Standard molecular biology protocols were adapted from ref. 8. The EGFP cDNA (CLONTECH) was fused to the C terminus of human CD4 by using an artificial *NotI* site. For alkaline phosphatase (AP) fusions, the SEAP cDNA (Tropix) was fused to human CD4 (partial cDNA encoding amino acids 380–458) with the stop codon replaced by a *Bsp120I* site. The linker between the SEAP and CD4 cDNAs reads ARRRKKRGLDV and encodes a furin cleavage site. Constructs were created by PCR and verified by sequencing. For transient

transfection, constructs were in pcDNA3 (Invitrogen). For the yeast two-hybrid system, constructs were cloned into vectors pGADT7 and pGBKT7 (CLONTECH). The exact amino acid sequence of the coiled-coil domain and surrounding linkers reads: GGGSGSRMKQIEDKLEEILSKLYHIENELAR-
IKKLLGERGGSGSAAA (9). To create stable HEK293 cells by Flp recombinase-mediated integration, constructs were in pcDNA5 (Invitrogen) and the manufacturer's protocol was followed exactly. Genomic DNA was prepared by using the DNeasy kit (Qiagen).

Generation of the Retroviral Libraries and Transduction of 3T3 Cells.

Tail libraries in the retroviral expression vector pLNCX (CLONTECH) were created by PCR using a primer directed against the 5' end of the EGFP cDNA and degenerate oligonucleotides (for the randomized positions each base was specified to be inserted with a 25% frequency; however, sequencing of control clones revealed that actual frequencies deviated from this; see Fig. 2) priming the 3' end of the CD4 cDNA. PCR products were digested with *NotI* and *PacI* and cloned into pLNCXhCD4 with a modified multiple cloning site. Following the last residue of EGFP, the exact amino acid sequence of the tails reads -TLA-SSLTFKXX-COOH, -TLASSLXXXKTN-COOH, and -TLASXXRXXSLS-COOH for the respective libraries.

Packaging cells (PT67, CLONTECH) were transiently transfected with the libraries using Fugene (Roche) in 80% confluent, 80-cm² tissue culture flasks (30 μ l of Fugene and 10 μ g of library DNA per flask; 1 flask for the -KKXX-COOH library, 2 flasks for the -XXXKTN-COOH library, and 6 flasks for the -XXRXXSLS-COOH library). After 36 h the viral supernatant was harvested and NIH 3T3 cells were infected at 70% confluency in 80-cm² tissue culture flasks. The cells were incubated in a 1:3 dilution of the viral supernatant containing polybrene at a final concentration of 8 μ g/ml for 10 h. After an additional 72 h in regular medium, cells were passaged and subjected to G418 selection (0.5 mg/ml) for 1 week before sorting. Infection rates under these conditions were determined by flow cytometry and ranged between 0.5% and 2%. From this we estimate the number of independent integrations to range from 50,000 to 100,000 for the respective libraries.

Flow Cytometry. Cells were detached by using 0.05% Trypsin in Versene buffer, harvested by centrifugation, and resuspended in 1 ml of medium per 10⁶ cells. PE-conjugated anti-hCD4 monoclonal antibody Q4120 (Sigma) was added at a dilution of 1:10 and staining was performed for 30 min on ice. Cells were washed once in medium containing 0.2% calf serum and 5% PBS-based

Abbreviations: AP, alkaline phosphatase; ER, endoplasmic reticulum; GFP, green fluorescent protein; TGN, trans-Golgi network.

*To whom reprint requests should be addressed: Zentrum für Molekulare Biologie (ZMBH), Universität Heidelberg, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany. E-mail: b.schwappach@zmbh.uni-heidelberg.de.

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.

cell dissociation buffer (Life Technologies), and resuspended in the same medium at 10^6 cells/ml. To assess viability, propidium iodide was added to a final concentration of $1 \mu\text{g/ml}$ immediately before sorting. Sorting was performed on a FACS Vantage SE cell sorter (Becton Dickinson). The 488 nm line of an Enterprise Laser (Innova) and the following filter sets were used: 530/30 bandpass (EGFP), 575/26 bandpass (PE), and 630/22 bandpass (PI). Stained, untransduced 3T3 cells and 3T3 cells transiently transfected with EGFP were used to set the compensation parameters.

Quantitative TGN Processing and Surface Expression Assays. For transient transfection, COS7 cells were plated in 35-mm tissue culture dishes (surface protein assay) or 24-well tissue culture plates (TGN processing assay) at 80% confluency. Cells were transfected with Fugene (2 μg of DNA and 5 μl of Fugene per 35-mm dish; 1 μg and 2.5 μl of Fugene per well of 24-well plate). Assays were performed 24 h after transfection. For the surface labeling assay, cells were fixed with 4% formaldehyde in PBS (30 min), blocked in PBS containing 1% goat serum (30 min), and labeled with mouse monoclonal antibody mAb 1779 (1:1000 dilution; Chemicon) (60 min) and goat anti-mouse horseradish peroxidase-conjugated secondary antibody (1:1000 dilution; Jackson) for 20 min. Chemiluminescence of the whole 35-mm dish was quantified in a TD20/20 luminometer (Turner Designs) after 15-s incubation in Power Signal Elisa solution (Pierce). Extensive washing was necessary to obtain a good signal-to-noise ratio; all incubations were performed at room temperature. For each data point, two to three dishes were averaged. For TGN processing assays, 24 h after transfection medium was removed from wells and heat-inactivated at 65°C for 1 h, and 4 μl was added to 60 μl of a one-step chemiluminescent assay solution (Tropix) and then incubated for 9 min. Chemiluminescent signal was measured in a TD-20/20 luminometer (Turner Designs). Data points are averages of three wells transfected with the same construct. For both assays, all constructs were tested in at least two independent experiments. In each experiment, a transfection with pcDNA3 vector was included to determine background as well as CD4-GFP AATN-COOH or AP-CD4-GFP AATN to allow for normalization between experiments.

Immunofluorescence. For the analysis of green fluorescent protein (GFP) staining patterns and immunofluorescence, COS7 cells were grown on glass chamber slides and transfected with Fugene. All cells were fixed in 4% formaldehyde in PBS for 30 min. GFP staining patterns were analyzed after fixation. For immunofluorescence, cells were incubated in 5% goat serum in PBS with 0.1% Triton X-100 (30 min), and labeled with primary antibody for 1 h and secondary antibody for 30 min (all steps at room temperature). All primary antibodies (rat anti-Grp74 (Stressgen), mouse anti-SRP54 (Transduction Labs), mouse anti-p115 (Transduction Labs), and mouse anti- γ -Adaptin (Sigma) were used at a 1:200 dilution. Secondary antibodies [Cy3-conjugated donkey anti-rat antibody and goat anti-mouse antibody (Jackson)] were used at a 1:250 dilution. Microscopy was performed with a Bio-Rad confocal microscope.

Yeast Two-Hybrid Assay. All yeast protocols were from ref. 8. Strain AH109 (CLONTECH) was transformed with the constructs indicated in the figure and plated on dropout medium lacking tryptophane and leucine to recover transformants. Individual colonies were streaked on dropout medium lacking tryptophane, leucine, histidine, and adenine to select for interacting pairs of DNA binding and activation domain fusion proteins.

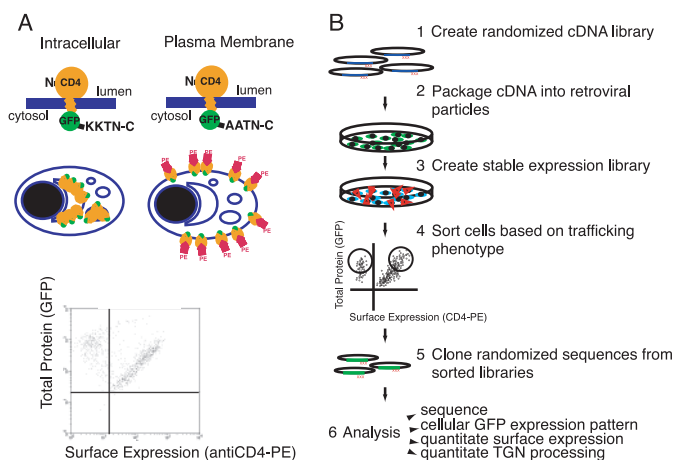


Fig. 1. Strategy for screening combinatorial libraries of sorting signal variants in intact mammalian cells. (A) Schematic depiction of the CD4-GFP reporter construct. GFP fluorescence is used to assess total reporter protein expression for individual cells, a phycoerythrin (PE) conjugated anti-CD4 antibody is used to measure CD4-GFP surface levels. Stable HEK293 cell lines expressing the reporter with an ER retention/retrieval sequence or an inactive version of the signal were created by Flp recombinase-mediated integration into the same genomic locus. Cells from the two lines were mixed, stained with the anti-CD4-PE antibody, and analyzed by flow cytometry (1,000 cells shown). Two populations reflecting the two trafficking phenotypes can readily be distinguished. (B) Flow chart detailing the individual steps of the screening strategy.

Results

Overview of Reporter Constructs and Combinatorial Library Screening in Mammalian Cells. To implement our retroviral-based screening strategy, we first developed a flow cytometry assay to measure ER retention of a reporter protein. To do this, we took advantage of the fact that strong ER localization results in reduced surface protein levels. However, this approach suffers from the problem that surface protein density is affected by the total protein expression level. To address this problem, we created a CD4 reporter protein with GFP fused to its C terminus (Fig. 1A). Using this reporter construct, intracellular localization can be measured as the ratio of surface protein (fluorescent antibody labeling of the extracellular domain of CD4) to total protein (GFP fluorescence), the value of which is independent of expression level. To test this approach, cells stably expressing a CD4 construct containing a C-terminal-KKXX-ER localization signal (from Wbp1p, refs. 10 and 11) were mixed with cells expressing CD4-GFP-AAXX which localizes to the plasma membrane (Fig. 1A). As predicted, two distinct populations were observed (Fig. 1A). Therefore, the CD4-GFP reporter construct can be used to distinguish ER retained proteins from ones localized to the plasma membrane.

Fig. 1B summarizes the overall procedure for generating and analyzing libraries of randomized signal variants. First, we created randomized libraries of sequences fused to CD4-GFP that were subcloned into a retroviral expression vector and used to transfect a packaging cell line. Retroviral particles in the supernatant were used to infect 3T3 cells. Low viral titers (<2% infection efficiency) ensured that each infected cell contained only one integration. Stably infected cells were sorted by flow cytometry for surface or intracellular localization. To identify tail sequences conferring different trafficking effects, we isolated genomic DNA from sorted cells and used PCR to amplify the randomized tail sequence. The randomized tails from sorted pools were then subcloned as fusions to CD4 in a transient expression vector. Clones were then sequenced and analyzed individually by transient transfection in COS-7 cells.

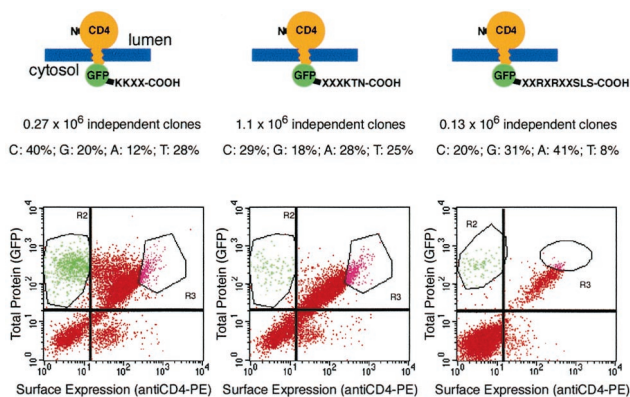


Fig. 2. Libraries of combinatorial variants of di-lysine and arginine-based ER retention/retrieval signals. Schematic depiction of the three libraries analyzed, the randomized positions are indicated in the C-terminal tail. The frequencies of the individual bases were determined by sequencing six to eight randomly picked clones from the unselected libraries. Populations of 3T3 cells infected with each library were sorted on a FACS Vantage SE cell sorter. Gates R2 and R3 were chosen to separate cells according to the two trafficking phenotypes of the reporter protein (retained or expressed on the surface). The plots shown are gated on viable cells as determined by propidium iodide staining (gate R1, not shown).

Application of Combinatorial Library Screen to ER Localization Signals.

We used the method outlined above to screen three libraries containing combinatorial variants of ER retention/retrieval

signals (Fig. 2). Statistics for each library and flow cytometry plots with the respective sorting windows are shown in Fig. 2. Frequency plots (Fig. 3) for each randomized position of the -KKXX-COOH and -XXXKTN-COOH libraries identify amino acid side chains favored or disfavored at the respective position (Fig. 3A). Commonly occurring residues that did not influence the efficacy of the signal appear on the diagonal (e.g., valine, leucine, isoleucine). For the -2 position of the -KKXX-COOH library arginine, threonine, phenylalanine, and particularly tyrosine were clearly favored, whereas proline or alanine were frequently found in nonfunctional tails (Fig. 3A). Stop codons were only found in the pool of tails conferring surface expression. This finding is consistent with an absolute requirement for two residues between the di-lysines and the C terminus (12). Leucine was the most common amino acid selected at the -1 position in tails conferring intracellular retention. Interestingly, 8 of 12 possible codon combinations for -KKYL-COOH were selected in the intracellular retained pool, suggesting that this is an especially potent combination. Analysis of the -XXXKTN-COOH library revealed a strong preference for lysine in the -4 position (Fig. 3B), as expected. However, lysine was not favored to the same extent in the -5 position, suggesting that most -XXKXXX-COOH sequences do not function as strong ER retention/retrieval signals (Fig. 3B).

Fewer unique sequences were isolated from the -XXRXRXSLS-COOH library. The results are summarized as a table of selected sequences (Fig. 3C). One striking feature of the sequences conferring intracellular retention is the frequency of

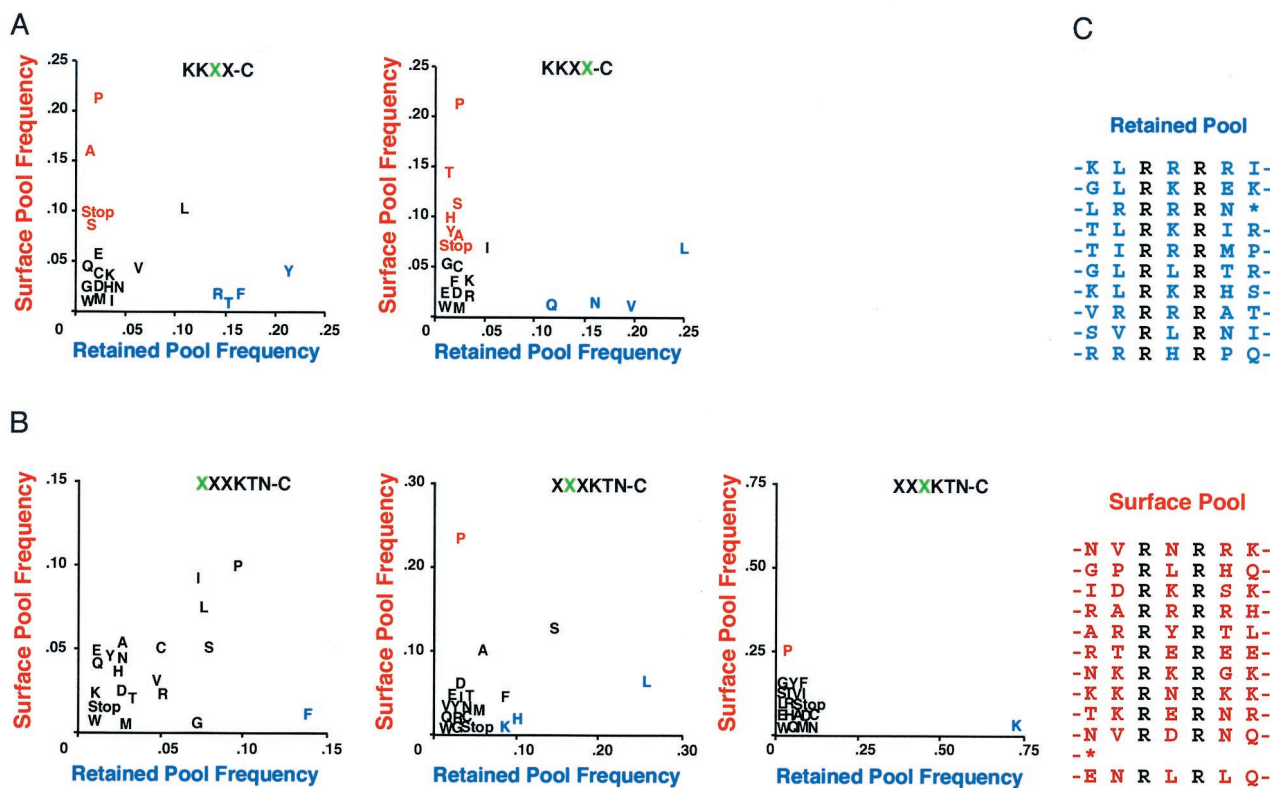


Fig. 3. Amino acid frequency at flanking positions of di-lysine and arginine-based signals. (A) Frequency plots for each randomized position in the di-lysine signal libraries. Unique sequences identified for both libraries were analyzed with respect to the amino acid frequency at each randomized position in the pool of intracellularly retained (x axis) or surface-expressed (y axis) reporter proteins. The number of unique sequences were 56 (-KKXX-C, retained), 48 (-KKXX-C, surface), 39 (-XXXKTN-C, retained) and 39 (-XXXKTN, surface). The position analyzed in the respective plot is indicated in green. Amino acids favored in the surface pool are indicated in red, whereas amino acids favored in the retained pool are shown in blue. Amino acids shown in black were either rare or equally abundant in both pools. (B) Table of clones selected from the library of arginine-based signal variants. Blue randomized positions indicated ER localization, whereas red randomized positions indicate surface expression. Tables 1 and 2 (which are published as supplementary material on the PNAS web site, www.pnas.org) list the original sequences obtained from the screen.

arginines in addition to the two fixed positions. Hydrophobic residues were clearly favored in the position immediately preceding the first arginine in the cluster. Negatively charged side chains as well as asparagine were strongly enriched in the pool of nonfunctional tail sequences.

To verify that our FACS analysis separated signal variants with different trafficking properties we tested whether individual tail sequences functioned as ER retention/retrieval signals by quantifying their effect on TGN processing and plasma membrane protein levels. To assess TGN processing, we created a reporter construct in which secreted placental AP was fused to the transmembrane domain of CD4 with a linker containing a furin protease site. Furin protease activity is restricted to the TGN and later compartments (see ref. 13 for a recent review). Therefore, AP is only cleaved and secreted when the reporter protein exits the ER and passes through the TGN. We fused GFP-tail sequences from our screen to the C terminus of the AP-CD4 reporter construct, and measured AP secretion with a chemiluminescent enzyme assay after transient transfection. To quantify plasma membrane protein levels, we transiently transfected COS-7 cells and labeled the CD4-GFP reporter protein on the surface of nonpermeabilized cells by using anti-CD4 primary and a horseradish peroxidase-conjugated secondary antibody. Captured horseradish peroxidase on the surface of cells was measured by single dish luminometry.

Fig. 4 summarizes TGN processing and surface protein levels for selected clones from all three libraries. Both di-lysine signals and arginine-based signals displayed a continuous spectrum of strong, intermediate, and nonfunctional signals. Clones isolated from the intracellular pool showed low TGN processing and surface expression, indicating that the FACS analysis accurately distinguishes between signals with different trafficking phenotypes.

We investigated how the quantitative differences in access to the TGN and cell surface (Fig. 4) translate into differences in cellular staining patterns (Fig. 5). CD4-GFP staining patterns were classified by using the following criteria: ER [perinuclear and reticulate staining pattern; confirmed by colocalization with SRP54 and GRP74 (14); Fig. 5A and data not shown], Golgi [juxtannuclear dotlike structure; confirmed by colocalization with p115 (15); Fig. 5A and data not shown], TGN (juxtannuclear more confined dotlike structure; confirmed by colocalization with γ -adaptin; Fig. 5A and data not shown), and cell surface (as assessed by surface staining and flow cytometry or tissue culture dish luminometry; Figs. 2 and 4). Fig. 5B shows a series of GFP staining patterns for related tails varied at only one position (-KKYX-COOH, -KXXN-COOH, -KKXL-COOH), yet each series covers the entire spectrum from predominately ER to predominately surface localization. Similarly, a series of clones containing the same central RKR motif gave rise to a whole range of steady-state staining patterns depending on the flanking residues.

Yeast Two-Hybrid Analysis of ER Trafficking Sequences. Di-lysine ER retention/retrieval signals have been shown to bind to the COPI vesicle coat (16, 17). However, substantial controversy exists as to which COPI subunit binds the di-lysine sequence (18, 19). We decided to examine this question using strong signal variants identified in our screen in a two-hybrid assay against all seven subunits of the COPI coat (20). To increase the effective local concentration of tail sequences, we fused them to a GAL4 DNA-binding or activation domain containing a sequence that mediates assembly of parallel four-stranded coiled-coil structures (9). We observed strong reporter gene activity when we cotransformed the GAL4 DNA-binding and activation domains in which both contained the coiled-coil sequence, but not when only one contained the coiled-coil sequence (Fig. 6A, right column). This result confirms the assembly of our two-hybrid constructs.

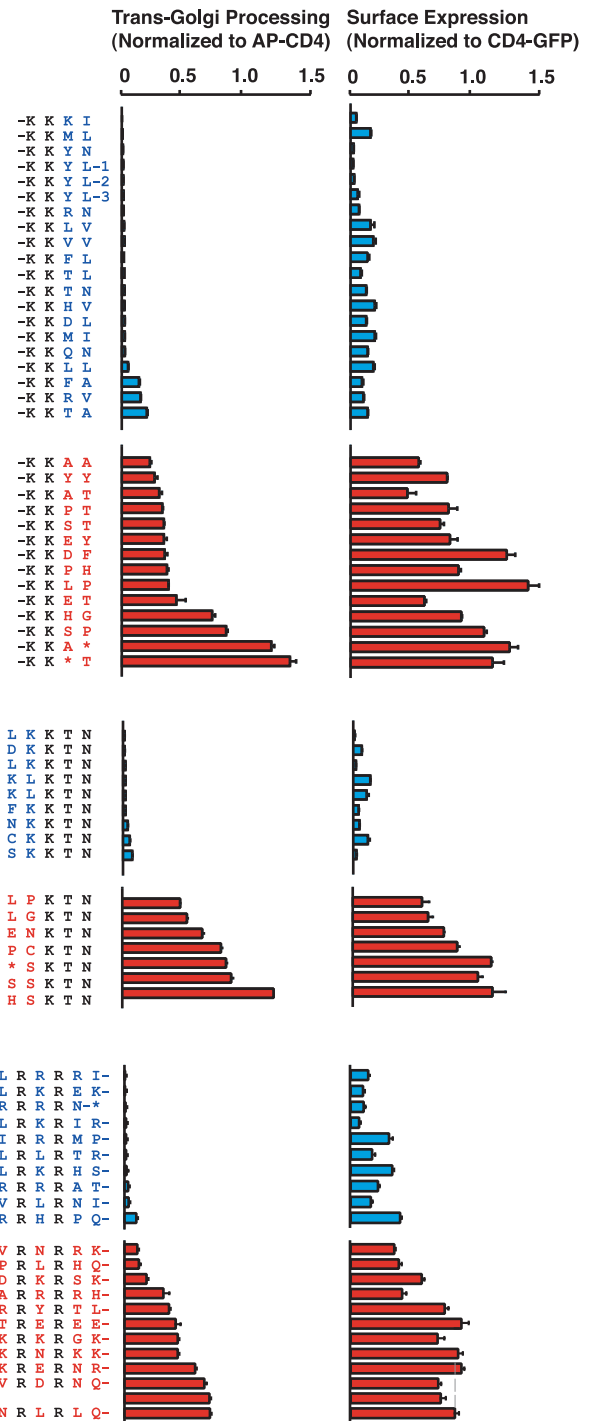


Fig. 4. Quantitative assays for TGN processing and surface expression of selected library clones. Fixed positions are shown in black, sequences of tails conferring ER retention/retrieval are shown in blue, and sequences of tails allowing for surface expression are shown in red. TGN processing was assessed by measuring alkaline phosphatase activity in the supernatant of transiently transfected COS-7 cells (left column). Surface expression was quantified by antibody staining and luminometry (right column).

First, we tested a strong version of the -KKXX-COOH signal (-KKYL-COOH) and an inactive version (-KK-COOH) for interaction with each of the seven subunits of the COPI coat (20). Coexpression of -KKYL-COOH, but not -KK-COOH, with the α -COPI subunit resulted in strong induction of the reporter genes (Fig. 6A). No interaction was detected between -KKYL-

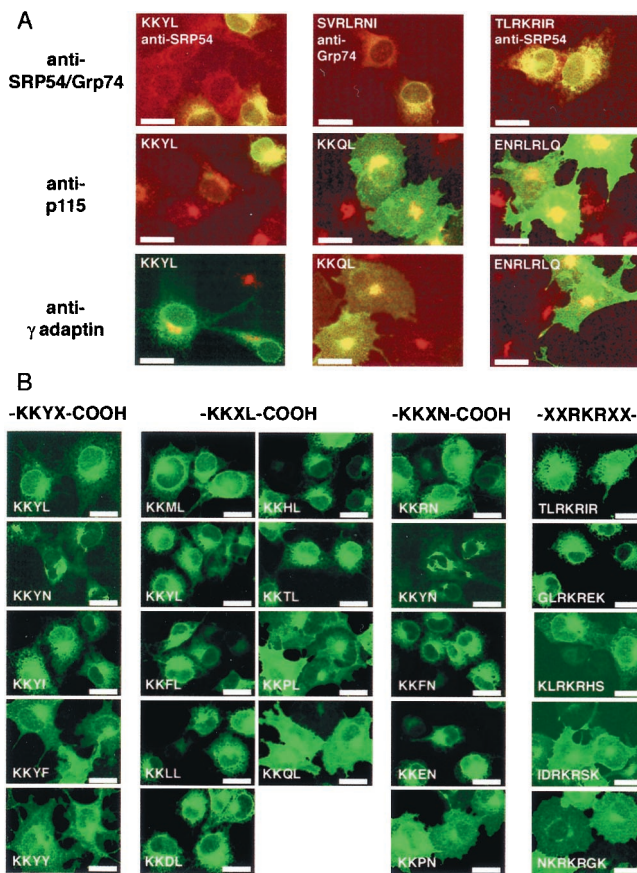


Fig. 5. Cellular staining patterns of selected signal variants. (A) Costaining of selected clones with antibodies directed against marker proteins for the ER (anti-SRP54/anti-Grp75), Golgi (anti-p115), and the TGN (anti- γ -adaptin). Fluorescence of CD4-GFP fusion proteins is merged with red fluorescence of stained marker proteins (Cy3-conjugated secondary antibodies). (B) GFP staining patterns for di-lysine signals with only the -1 or -2 position varied and tails containing an RKR signal with different flanking residues. Each series covers the whole range from mainly ER to surface localization. (Bar = 25 μ m.)

COOH and other COPI subunits (however, as previously reported (20), expression of the β - and ζ -COPI subunits as DNA-binding domain fusions resulted in activation independent of coexpression with the GAL4 activation domain). The tail sequence containing an arginine-based signal did not specifically interact with any of the seven COPI subunits tested (Fig. 6B). Sequences that we characterized as strong ER retention/retrieval signals (-KKML-COOH and -KKLV-COOH) were also found to interact with α -COP as assayed by the two-hybrid system (Fig. 6C). Intermediate or inactive sequences (-KKTA-COOH, -KKFA-COOH, -KKAA-COOH, -KKPH-COOH, and -KKSP-COOH) did not exhibit interaction (Fig. 6C). Interestingly, the tail ending in -KKYY-COOH interacted with α -COP as assayed by the two-hybrid system but was expressed on the surface in COS-7 cells. These interactions were specific to the di-lysine signal, because changing -KKYL-COOH, -KKML-COOH, and -KKYY-COOH to -AAYL-COOH, -AAML-COOH, -AAYY-COOH abolished interactions with α -COP (Fig. 6C). The finding that α -COP interacts with a subset of di-lysine containing sequences that we have identified as strong ER retention/retrieval signals suggests that α -COP is the receptor for di-lysine trafficking signals. Furthermore, under the same two-hybrid conditions, no interaction was observed between an active arginine-based sequence and any of the COPI coat subunits (Fig. 6B), raising the possibility that di-arginine signals use a different mechanism than di-lysine signals.

Discussion

The screening method described here should be generally applicable to the study of sorting signals in intact mammalian cells. Although our initial screens focused on ER localization signals, this method could be adapted to the analysis of other known signals for ER export, endocytosis, lysosomal targeting, or degradation. In contrast to biochemical or two-hybrid analysis of interactions between short peptide sequences and adaptor or coat proteins, the screening approach described here assays the combined effect of many molecular interactions (even very low affinity ones) in the intact cellular environment. Our approach can provide useful information about the sequence consensus for a trafficking signal even if the receptor for the motif is not known.

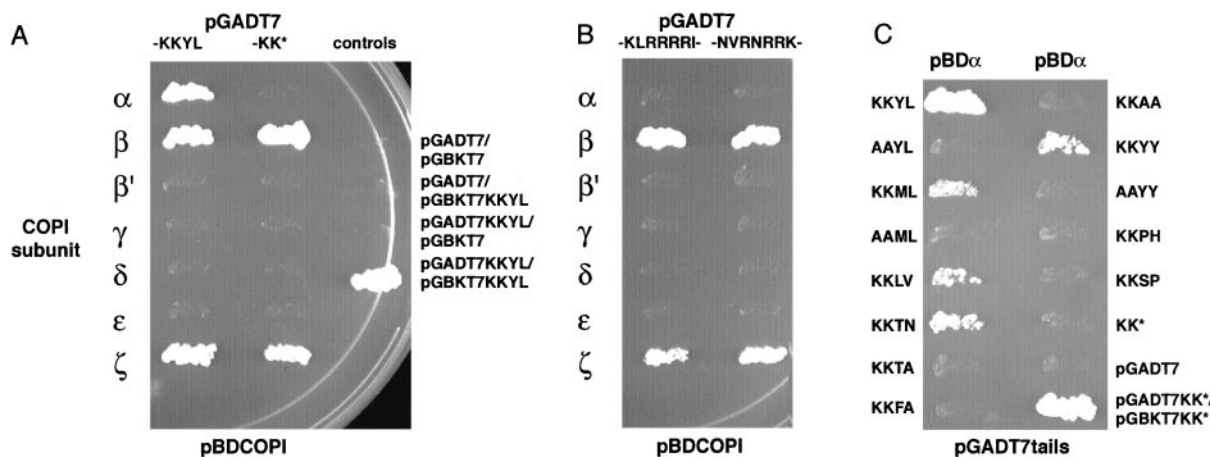


Fig. 6. Yeast two-hybrid analysis of interaction between selected tails from the screen and subunits of the COPI vesicle coat. By using an active and inactive version of the di-lysine or the arginine-containing tail as a bait, all seven subunits of the COPI vesicle coat were tested for interaction (A and B). Indicated constructs were cotransformed and colonies were streaked on selective medium (dropout without histidine and adenine) after 3 days. Pictures were taken after 4 days of growth on selective medium. Specific interaction with the α -COP subunit was observed for the -KKYL-COOH (not the -KK-COOH) tail (A), but not for the arginine-containing tail (B). Bait constructs containing different versions of the di-lysine signal were tested against the α -COP subunit (C). Strong versions of the signal isolated from the screen interacted with α -COP, whereas signals of intermediate strength or inactive versions gave no signal.

The results from our screen are generally compatible with previous studies of the -KKXX-COOH consensus (refs. 1, 12, 21, and 22 and references therein). One of the most surprising results is the finding that combinatorial variants of either lysine- or arginine-based ER retention/retrieval signals can generate a broad spectrum of trafficking signals with widely varying efficacy. For the ER-Golgi intermediate compartment marker ERGIC-53, the -1 and -2 position have previously been shown to influence the ER targeting capacity of the C-terminal di-lysine signal of ERGIC-53 (23), e.g., -KKAA was shown to confer ER localization much more efficiently than -KKFF (24). In the context of our reporter CD4-GFP, -KKAA was not the strongest possible version of the signal. The signal variants in our screen were fused to GFP, which is expected to weaken the strength of the trafficking signals by positioning them away from the membrane (25). It seems however likely that the relationship between the signal variants observed in our screen is consistent with previous data as -KKYY and -KKYF-bearing CD4-GFP reporter selected in our screen showed much stronger TGN processing and surface expression than the -KKAA variant. The result that ER retention/retrieval signals have variable efficacy raises the possibility that the strength of each trafficking signal can be finely tuned to achieve the appropriate steady-state distribution.

Many studies have documented the interaction between the COPI coat and di-lysine signals. It remains controversial, however, which subunit of the COPI coat binds the di-lysine signal. Site-directed photocrosslinking studies (19, 26) show selective crosslinking of di-lysine peptides to the COPI γ -subunit. However, binding experiments have pointed to a role for the α -COPI subunit (16). Genetic approaches have shown that mutations in α , β' , γ , δ , and ζ -COP subunits can interfere with -KKXX retrieval (17, 27). Furthermore, deletion of the N-terminal WD40 domain of the α -COP subunit yields a specifically retrieval-deficient COPI complex that seems to be fully functional otherwise (28). We observed that the strongest -KKXX-COOH signals identified in our screen selectively interacted with the α -subunit of the COPI vesicle coat in the yeast two-hybrid assay. Importantly, our ranking of di-lysine *in vivo* efficacy closely matched the results from the two-hybrid interaction assay. Therefore, α -COP preferentially interacts with a subset of

-KKXX-COOH variants that also confer a strong ER-localization phenotype. This close correlation argues in favor of α -COP as the *in vivo* receptor for di-lysine signals.

Arginine-based ER trafficking signals (1, 5) have recently been shown to play an important role in membrane protein quality control processes (4, 29, 30). In contrast to di-lysine signals which exhibit a strict spacing relative to the C terminus, arginine signals are found in a variety of sequence positions (cytosolic N terminus, intracellular loops, and cytosolic C terminus; refs. 4 and 5). Because of the high frequency of di-arginine sequences in the cytosolic domains of membrane proteins, the results from our screen should be useful for predicting the presence of functional di-arginine trafficking signals. Our screen demonstrates that clusters of arginines act as particularly strong ER localization signals (consistent with the sequence analyzed in ref. 5) and that a hydrophobic amino acid, in particular leucine, usually precedes the arginine cluster. These sequence preferences are observed in the arginine motifs found in an inwardly rectifying potassium channel (4), an ATP-binding cassette (ABC) protein (4), and in the GABA_B G-protein coupled receptor (30).

The screens described here demonstrate that subtle differences in the sequence context surrounding sorting signals can dramatically alter the steady-state distribution of a protein. Therefore, the predictive power of a consensus motif is greatly increased if the flanking positions are taken into consideration. Our screens show that many signal variants are nonfunctional and identify variants that are especially potent. This refined consensus information is likely to help evaluate the probability that the occurrence of a trafficking signal contributes to the trafficking behavior of the protein.

We thank many Jan laboratory members for stimulating discussions and comments on the manuscript. In particular, we thank Dan Minor for discussing multivalent yeast two-hybrid baits and for the cDNA encoding the four-stranded coiled-coil. We are most grateful to Dorian Faulstich and Felix Wieland for the generous gift of 14 constructs of COPI subunits in two-hybrid vectors. We are indebted to the neurogenetics sequencing core facility and the Howard Hughes sequencing facility at the University of California, San Francisco for excellent sequencing support. B.S. was supported by a Human Frontier Science Program postdoctoral fellowship. N.Z. is the recipient of a Howard Hughes predoctoral fellowship. L.Y.J. and Y.N.J. are Howard Hughes investigators.

- Teasdale, R. D. & Jackson, M. R. (1996) *Annu. Rev. Cell Dev. Biol.* **12**, 27–54.
- Kirchhausen, T. (1999) *Annu. Rev. Cell Dev. Biol.* **15**, 705–732.
- Bonifacino, J. S. & Dell'Angelica, E. C. (1999) *J. Cell Biol.* **145**, 923–926.
- Zerangue, N., Schwappach, B., Jan, Y. N. & Jan, L. Y. (1999) *Neuron* **22**, 537–548.
- Schutze, M. P., Peterson, P. A. & Jackson, M. R. (1994) *EMBO J.* **13**, 1696–1705.
- Ohno, H., Stewart, J., Fournier, M. C., Bosshart, H., Rhee, I., Miyatake, S., Saito, T., Gallusser, A., Kirchhausen, T. & Bonifacino, J. S. (1995) *Science* **269**, 1872–1875.
- Boll, W., Ohno, H., Songyang, Z., Rapoport, I., Cantley, L. C., Bonifacino, J. S. & Kirchhausen, T. (1996) *EMBO J.* **15**, 5789–5795.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1997) *Current Protocols in Molecular Biology* (Wiley, New York).
- Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993) *Science* **262**, 1401–1407.
- Townsend, F. M. & Pelham, H. R. (1994) *Eur. J. Cell Biol.* **64**, 211–216.
- Gaynor, E. C., te Heesen, S., Graham, T. R., Aebi, M. & Emr, S. D. (1994) *J. Cell Biol.* **127**, 653–665.
- Jackson, M. R., Nilsson, T. & Peterson, P. A. (1990) *EMBO J.* **9**, 3153–3162.
- Molloy, S. S., Anderson, E. D., Jean, F. & Thomas, G. (1999) *Trends Cell Biol.* **9**, 28–35.
- Rapiejko, P. J. & Gilmore, R. (1997) *Cell* **89**, 703–713.
- Sapperstein, S. K., Walter, D. M., Grosvenor, A. R., Heuser, J. E. & Waters, M. G. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 522–526.
- Letourneur, F., Gaynor, E. C., Hennecke, S., Démollière, C., Duden, R., Emr, S. D., Riezman, H. & Cosson, P. (1994) *Cell* **79**, 1199–1207.
- Cosson, P. & Letourneur, F. (1994) *Science* **263**, 1629–1631.
- Gaynor, E. C., Graham, T. R. & Emr, S. D. (1998) *Biochim. Biophys. Acta* **1404**, 33–51.
- Harter, C. & Wieland, F. T. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 11649–11654.
- Faulstich, D., Auerbach, S., Orci, L., Ravazzola, M., Wegchlingel, S., Lottspeich, F., Stenbeck, G., Harter, C., Wieland, F. T. & Tschochner, H. (1996) *J. Cell Biol.* **135**, 53–61.
- Jackson, M. R., Nilsson, T. & Peterson, P. A. (1993) *J. Cell Biol.* **121**, 317–333.
- Nilsson, T., Jackson, M. & Peterson, P. A. (1989) *Cell* **58**, 707–718.
- Itin, C., Schindler, R. & Hauri, H. P. (1995) *J. Cell Biol.* **131**, 57–67.
- Andersson, H., Kappeler, F. & Hauri, H. P. (1999) *J. Biol. Chem.* **274**, 15080–15084.
- Vincent, M. J., Martin, A. S. & Compans, R. W. (1998) *J. Biol. Chem.* **273**, 950–956.
- Harter, C., Pavel, J., Coccia, F., Draken, E., Wegchlingel, S., Tschochner, H. & Wieland, F. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 1902–1906.
- Cosson, P., Démollière, C., Hennecke, S., Duden, R. & Letourneur, F. (1996) *EMBO J.* **15**, 1792–1798.
- Eugster, A., Frigerio, G., Dale, M. & Duden, R. (2000) *EMBO J.* **19**, 3905–3917.
- Chang, X. B., Cui, L., Hou, Y. X., Jensen, T. J., Aleksandrov, A. A., Mengos, A. & Riordan, J. R. (1999) *Mol. Cell* **4**, 137–142.
- Margeta-Mitrovic, M., Jan, Y. N. & Jan, L. Y. (2000) *Neuron* **27**, 97–106.