

The Medium Subunits of Adaptor Complexes Recognize Distinct but Overlapping Sets of Tyrosine-based Sorting Signals*

(Received for publication, May 27, 1998)

Hiroshi Ohno‡, Ruben C. Aguilar§, David Yeh§, Daisuke Taura§, Takashi Saito‡, and Juan S. Bonifacino§¶

From The ‡Division of Molecular Genetics, Chiba University Graduate School of Medicine, 1-8-1 Inohana, Chuo-ku, Chiba 260-8670, Japan and the §Cell Biology and Metabolism Branch, NICHD, National Institutes of Health, Bethesda, Maryland 20892

Tyrosine-based sorting signals conforming to the motif YXXØ (Y is tyrosine, X is any amino acid, and Ø is an amino acid with a bulky hydrophobic side chain (leucine, isoleucine, phenylalanine, methionine, valine)) interact with the medium (μ) subunits of clathrin adaptor (AP) complexes. We have analyzed the selectivity of interaction between YXXØ signals and the μ 1, μ 2, and μ 3 (A or B) subunits of the AP-1, AP-2, and AP-3 complexes, respectively, by screening a combinatorial XXXYXXØ library using the yeast two-hybrid system. All the medium subunits were found to prefer proline at position Y+2, suggesting that YXXØ signals are stabilized by a bend in the polypeptide backbone. Other than for this common preference, each medium subunit favored specific sets of residues at the X and Ø positions; these preferences were consistent with the proposed roles of the different adaptor complexes in rapid endocytosis and lysosomal targeting. A considerable specificity overlap was also revealed by these analyses, suggesting that additional factors, such as the context of the signals, must be important determinants of recognition.

Targeting of integral membrane proteins to compartments of the endosomal-lysosomal system and to the basolateral plasma membrane of polarized epithelial cells is largely dependent upon sorting signals contained within the cytosolic domains of the proteins (reviewed in Refs. 1–3). The signals consist of short, degenerate sequences that conform to any of several consensus motifs. One of the most extensively characterized motifs contains a critical tyrosine residue within the canonical sequence YXXØ (Y is tyrosine, X is any amino acid, and Ø is an amino acid with a bulky hydrophobic side chain (leucine, isoleucine, phenylalanine, methionine, valine)) (Refs. 4–7; reviewed in Ref. 2). The activity of this type of sorting signal requires that the critical tyrosine be in an unphosphorylated state (8, 9). Tyrosine-based sorting signals conforming to the YXXØ motif have been shown to interact directly with the medium subunits μ 1, μ 2, and μ 3 (A or B) of the adaptor complexes AP-1, AP-2, and AP-3, respectively (8–17) (μ 3A and μ 3B are ubiquitous and neuronal-specific forms, respectively,

of the medium subunit of AP-3). YXXØ-adaptor interactions result in selective incorporation of the integral membrane proteins into coated vesicles that carry proteins to different destinations within the cell. AP-1 is responsible for the delivery of proteins from the *trans*-Golgi network (TGN)¹ to the endosomal-lysosomal system, whereas AP-2 mediates rapid internalization of endocytic receptors from the plasma membrane (1–3). Recent studies suggest that AP-3 is involved in an alternative pathway of protein transport from the TGN or endosomes to lysosomal compartments (18–22).

The involvement of the same type of YXXØ signal in protein sorting to different compartments raises a question of specificity. It has been proposed that YXXØ signals are not all functionally equivalent but may be recognized with a characteristic fine specificity by each adaptor complex (2, 8). Mutational analyses of tyrosine-based signals have suggested that this is indeed the case, as both the identity of the X and Ø residues and the position of the signal within the cytosolic domain are important determinants of interactions with different adaptor medium subunits (8). However, targeted mutagenesis has proven too limited an approach for defining the requirements for specific recognition of tyrosine-based signals by adaptor medium subunits. A potentially more powerful approach is the screening of combinatorial libraries. This method has permitted a detailed analysis of associations between phosphotyrosine-containing sequences and protein interaction modules such as SH2 and PTB domains (23, 24). In a previous study (11), we performed an analysis of μ 2 binding specificity by screening a combinatorial XXXX tetrapeptide library using the yeast two-hybrid system. All the tetrapeptides selected by μ 2 conformed to the consensus sequence YXXØ, with YXRL tetrapeptides exhibiting the highest apparent affinity (11). However, attempts to use the XXXX library to characterize the sequence preferences of μ 1, μ 3A, and μ 3B were unsuccessful because of the presence of unfavorable amino acid residues (*i.e.* two alanines) amino-terminal to the tyrosine residue (8).

We have now investigated the selectivity for interaction of tyrosine-based sorting signals with μ 1, μ 2, μ 3A, and μ 3B through screening of a new combinatorial XXXYXXØ library using the yeast two-hybrid system. The results of these analyses revealed that each medium subunit binds YXXØ signals with distinct sequence preferences, although there is also a considerable degree of specificity overlap.

EXPERIMENTAL PROCEDURES

Construction of a Combinatorial Library for Screening by the Yeast Two-hybrid System—A DNA fragment encoding the 33-amino acid cy-

* This work was supported in part by grants-in-aids for scientific research from the Ministry of Education, Science, and Culture of Japan (to H. O. and T. S.), the Naito Foundation (to H. O.), and the Uehara Memorial Foundation (to H. O.). 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: NICHD, National Institutes of Health, Bldg. 18T Rm. 101, 18 Library Dr., MSC 5430, Bethesda, MD 20892-5430. Tel.: 301-496-6368; Fax: 301-402-0078; E-mail: juan@helix.nih.gov.

¹ The abbreviations used are: TGN, *trans*-Golgi network; 3AT, 3-amino-1,2,4-triazole; SED, standard error of the difference.

tosolic tail of TGN38 (25) was engineered to delete codons for the last 12 amino acids and to add an *EagI* site by introduction of silent mutations in place of the codons for Arg²¹ and Pro²² (AGGCCA to CGGCCG). This construct was made by polymerase chain reaction using the following primers: CGGAATTCACAACAACGA-AAGATTAT and CCGCTC-GAGCGCCGCTAGTGACTTTGGATCT. The polymerase chain reaction product was ligated into the *EcoRI* and *SalI* sites of pGBT9(*TRP1*) (CLONTECH) to yield the plasmid pGBT9-TGNΔ-*EagI*. A random oligonucleotide library of sequence CCCGGCCGAAGNNNNNNNNNTACNNNNNNNT(C/G)TGACTGCAGTC was annealed via a palindromic sequence at the 3' end containing the restriction sites *EagI* and *PstI* (underlined). After conversion to double-stranded DNA by mutually primed synthesis, the oligonucleotide was digested with *EagI* and *PstI* and ligated into pGBT9-TGNΔ-*EagI* cut with *EagI* and *PstI*. Competent *Escherichia coli* cells (Max Efficiency DH5α, Life Technologies, Inc.) were transformed with the resulting plasmids to yield a combinatorial DNA library with 1.3×10^7 transformants. The amino acid sequence encoded by the resulting clones was HNKRKIIAFALEGKRSKVTRRP-KXXXYYXØ. The critical tyrosine residue was fixed because it has been shown that tyrosine is essential for interaction with the adaptor medium subunits (8, 10). Fixing the tyrosine residue had the added advantage of reducing the complexity of the library. For the same reasons, the last amino acid was restricted to Leu, Ile, Phe, Met, or Val. The theoretical complexity of the library was 8.6×10^9 different amino acid sequences. Ten library clones were selected at random and their nucleotide sequences determined. The frequencies of the four nucleotides at the N positions of this random sample were as follows: G, 20; A, 40; T, 75; and C, 25. Based on these data, the expected frequencies for each amino acid at the X positions were as follows: Ala and Asp, 1.95%; Cys, 3.66%; Glu, 1.17%; Phe, 13.73%; Gly, 1.56%; His and Pro, 2.44%; Ile, 10.25%; Lys, 2.34%; Leu, 15.56%; Met and Gln, 1.46%; Asn and Thr, 3.91%; Arg, 3.13%; Ser, 9.28%; Val, 5.86%; Trp, 0.73%; Tyr, 7.32%; and stop codons, 5.86%. The expected frequencies at position Y+3 (Ø) were as follows: Phe, 26.04%; Ile, 13.89%; Leu, 36.46%; Met, 11.11%; and Val, 12.50%.

Two-hybrid Screening of the Combinatorial Library—The yeast strain HF7c (MATα) (CLONTECH) was first transformed with GAL4ad-μ1, -μ2, -μ3A, or -μ3B (all in the pACTII(*LEU2*) plasmid) and plated onto yeast dropout agar plates lacking leucine as described in the protocol for the MATCHMAKER two-hybrid system (CLONTECH). The μ2 construct used in the initial screening of the library was GAL4ad-3M2 (10). This construct lacks four nucleotides at the 3' end of the coding sequence and thus displays a weaker affinity for tyrosine-based sorting signals. Use of this weakened μ2 construct resulted in the selection of only those signals that displayed the strongest affinity for μ2. Subsequent analyses of interactions of all of the isolated signals with μ2 were done using GAL4ad-μ2^{+E5} (26) encoding full-length μ2; this construct exhibits stronger binding affinity for YXXØ signals. Transformants containing the different medium subunit constructs were retransformed with the combinatorial DNA library and selected on plates lacking leucine and tryptophan for selection of cotransformants and histidine for selection of interacting clones; colonies that grew on these plates were tested for β-galactosidase activity. Colonies expressing β-galactosidase were cultured in dropout medium containing leucine but lacking tryptophan to obtain cells carrying only the library plasmid and not the medium subunit plasmid. The resulting cells were then mated with the yeast strain Y187 (MATα) transformed with each of the GAL4ad-μ constructs or with pTD1-1 (SV40 large-T antigen cDNA in pACTII; negative control for histidine auxotrophy and β-galactosidase activity) to test the binding specificity of library clones. Quantitative assays for growth in the presence of varying concentrations of 3-amino-1,2,4-triazole (3AT, Fluka Chemie AG, Buchs, Switzerland) and β-galactosidase chemiluminescence assays were performed as described (14).

Statistical Analyses—Based on the sequences selected by each adaptor medium subunit from the combinatorial library, an experimental (observed) frequency was calculated for each residue at each position of the XXXYXXØ sequence. Preferences were evaluated by calculating the difference between the observed and the expected frequencies (ΔF) divided by the standard error of the difference (SED). SED values were calculated according to the following error propagation formula,

$$SED = (SEF_o^2 + SEF_e^2)^{1/2} \quad (\text{Eq. 1})$$

in which SEF represents the standard error of the frequencies (O, observed; E, expected) (27). SEF was calculated according to the formula

$$SEF = [f(1 - f)/n]^{1/2} \quad (\text{Eq. 2})$$

in which *f* is the residue frequency and *n* is the number of sequences analyzed (27). We considered to be significantly different from 0 (random) any ΔF/SED value above 2 (*i.e.* favored) or below -2 (*i.e.* disfavored). In some cases, we lowered the stringency of the statistical analyses and considered as significant values of ΔF/SED higher than 1 or lower than -1.

RESULTS

Screening of an XXXYXXØ Combinatorial Library Using the Yeast Two-hybrid System—A combinatorial plasmid library encoding the sequence XXXYXXØ fused to the carboxyl terminus of GAL4bd was coexpressed with GAL4ad-μ1, -μ2, -μ3A, or -μ3B plasmid constructs into the appropriate yeast strain. Library clones that conferred the ability to grow in medium lacking histidine and that exhibited β-galactosidase activity when expressed in combination with either of the medium subunits were isolated for further analyses. The total numbers of clones isolated for each medium subunit were as follows: 36 for μ1, 33 for μ2, 52 for μ3A, and 14 for μ3B.

The amino acid sequences of the isolated XXXYXXØ clones, as inferred from DNA sequencing, are listed in Fig. 1. Each of these clones was tested for interaction with μ1, μ2, μ3A, or μ3B using yeast two-hybrid plate assays (Fig. 1). The frequencies with which specific amino acid residues appeared at each of the X and Ø positions (*i.e.* observed frequencies) were compared to the random frequencies predicted from the nucleotide composition of the library (*i.e.* expected frequencies) as described under "Experimental Procedures." The results of these analyses for each of the medium subunits are shown in Fig. 2. Positive or negative ΔF/SED values correspond to residues that were favored or disfavored, respectively. Residues with ΔF/SED values >2 or <-2, considered to be the most significant, are shown in Table I. Also shown in Table I are some residues with values of $1 < \Delta F/SED < 2$ or $-1 > \Delta F/SED > -2$, which were judged to be significant with a lower level of stringency. The most salient features of the signals revealed by these analyses are discussed in the following sections.

Degeneracy of YXXØ Signals—Previous studies of the function of YXXØ signals in intact cells established that the signals are highly degenerate (Refs. 4–7, reviewed in Ref. 2; see also Table II). The combinatorial analyses reported here confirmed this characteristic of YXXØ signals at the level of interactions with the adaptor medium subunits. The results showed that there was no absolute requirement for the presence of specific residues at any of the X or Ø positions. This contrasted with the critical tyrosine (Tyr) residue that could not be substituted by any other residue without a dramatic decrease in sorting activity (2, 4–7) and binding affinity for the medium subunits (8–11). Also of note is the fact that none of the sequences isolated in the combinatorial screens conformed to the NPXY motif, another consensus sequence for tyrosine-based endocytic signals (28). This could be due to the presence of unfavorable residues upstream of the first X residue of the library or to an intrinsically lower binding affinity of NPXY motifs for the medium subunits. Another possibility is that NPXY signals are not recognized by the adaptor medium subunits but by other components of clathrin coats. Although many different residues were tolerated at the X and Ø positions of YXXØ signals, each medium subunit exhibited characteristic preferences for certain sets of residues, as discussed below.

Preferences Common to All Medium Subunits—All medium subunits displayed a clear preference for proline at Y+2 (Fig. 2 and Table I). In addition, μ1, μ3A, and μ3B, but not μ2, preferred arginine at Y-3 (Fig. 2 and Table I). In general, residues with hydrophobic side chains were disfavored at the X positions (Table I).

μ1 Preferences—μ1 was found to prefer arginine at Y-3, serine at Y-2, leucine or aspartic acid at Y-1, proline at Y+2,

Sequences selected by:

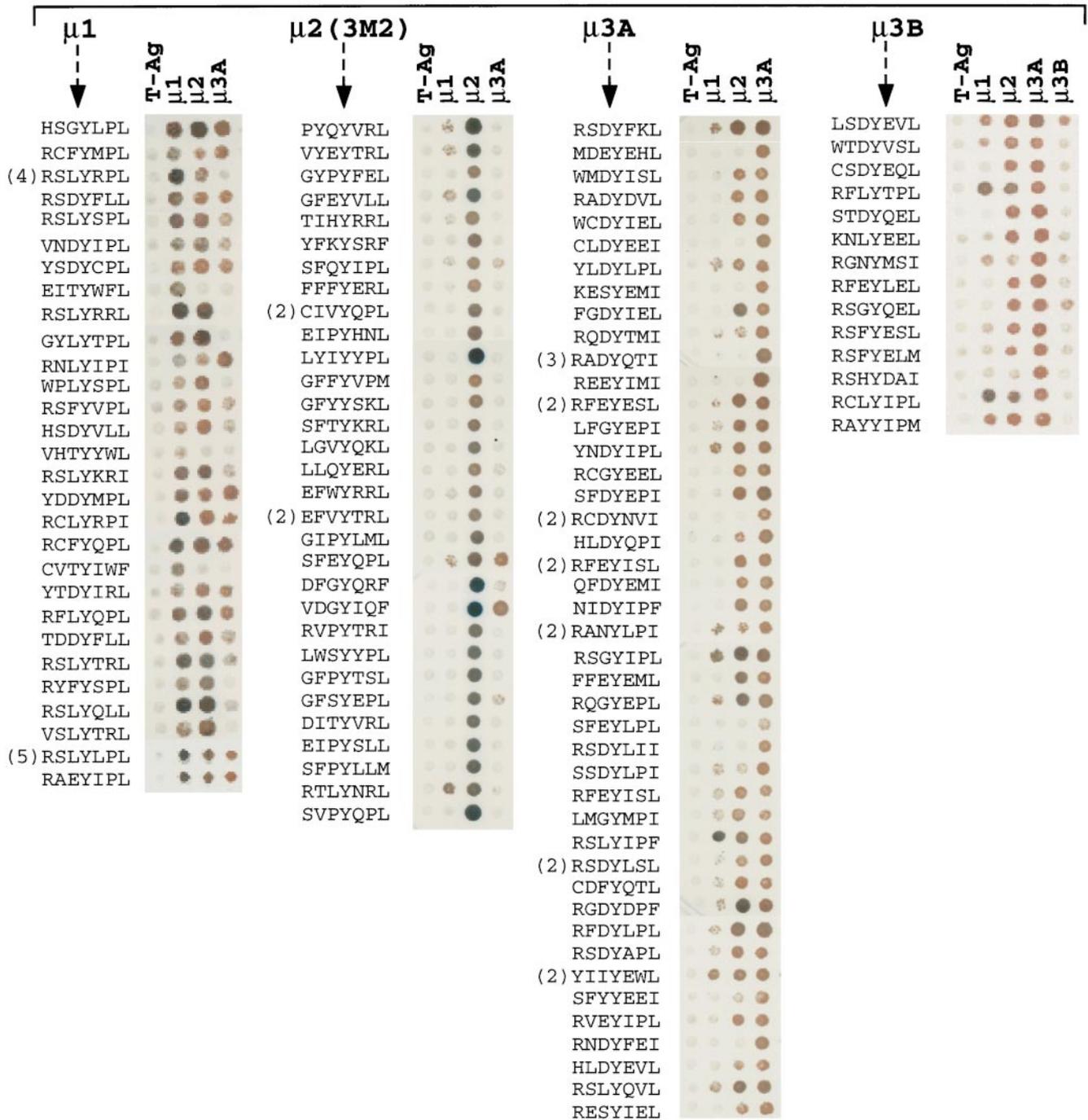


FIG. 1. Sequences and specificity of interactions of XXXYXXØ clones selected by $\mu 1$, $\mu 2$, $\mu 3A$, or $\mu 3B$. See "Experimental Procedures" for details. The number of clones having the same sequence are indicated in parentheses. To analyze the specificity of interaction of XXXYXXØ clones with the different medium subunits, the yeast strain HF7c (MAT α) expressing each GAL4bd-XXXYXXØ clone was mated with the Y187 strain (MAT α) expressing GAL4ad- $\mu 1$, - $\mu 2$, - $\mu 3A$, - $\mu 3B$, or -SV40 large-T antigen (T-Ag). Yeast transformants were plated onto dropout agar plates lacking leucine, tryptophan, and histidine supplemented with X-gal plus isopropylthiogalactoside. This allowed simultaneous determination of growth in the absence of histidine and β -galactosidase activity. Clones were considered to be positive if they were able to grow on the -His plates. Expression of β -galactosidase activity (dark blue colonies) identified the strongest interactions.

and leucine at Y+3 (Ø) (Fig. 2 and Table I). Disfavored residues were phenylalanine, leucine, or isoleucine at Y-3, leucine at Y-2, isoleucine at Y-1 and Y+2, and phenylalanine, methionine, or valine at Y+3 (Ø) (Fig. 2 and Table I). Interestingly, the sequences RSLYRPL and RSLYLPL were independently isolated four and five times, respectively (Fig. 1). The high frequency of these sequences may be due to the fact that they

contain all the residues identified as favorable for interactions with $\mu 1$ (Table I).

$\mu 2$ Preferences—Preferences for $\mu 2$ were glycine at Y-3, phenylalanine at Y-2, proline at Y-1, proline or arginine at Y+2, and leucine at Y+3 (Ø) (Fig. 2 and Table I). The preferences for arginine at Y+2 and leucine at Y+3 (Ø) are in agreement with the results from our previous screening of an XXXX

TABLE I
Summary of preferences

This table summarizes some of the results shown in Fig. 2. Residues were considered to be significantly favored or disfavored when their $\Delta F/SED$ values were >2 or <-2 , respectively. Residues in parentheses were considered to be favored or disfavored with a lower level of statistical significance ($\Delta F/SED > 1$ or $\Delta F/SED < -1$). Residues at the Y+3 (\emptyset) position were judged to be favored or disfavored relative to other residues with bulky hydrophobic side chains (leucine, isoleucine, phenylalanine, methionine, valine).

μ	Preference	Y-3	Y-2	Y-1	Y	Y+1	Y+2	Y+3 (\emptyset)
$\mu 1$	Favored	R	S	L, D			P	L
	Disfavored	F, L, I	L	I			I	F, M, V
$\mu 2$	Favored	G	F	P			P, R	L
	Disfavored						F	V
$\mu 3A$	Favored	R		D, E, (G)		E	P	I
	Disfavored	I				S	F, L	F, M, V
$\mu 3B$	Favored	R	(S)	(D)		E	(P), (E)	(L)
	Disfavored							F

TABLE II
YXX \emptyset -type signals involved in internalization and lysosomal targeting

Sequences were obtained from either the GenBank or Swissprot databases. The critical tyrosine is indicated in bold type. Three residues on either side of the critical tyrosine are shown for comparison with the YXX \emptyset sequences isolated in the combinatorial screens. Some of the internalization signals may also participate in transport to late endosomal compartments, the TGN, or the basolateral plasma membrane of polarized epithelial cells.

Proteins	Tyrosine-based signals
Internalization signals	
Transferrin receptor (human)	PLSYTRF
Asialoglycoprotein receptor H1 (human)	TKEYQDL
CI Man 6-P receptor (bovine)	SYKYSKV
CD Man 6-P receptor (bovine)	PAAYRGI
CD1b (human)	RRSYQNI
Poly-Ig receptor (mouse)	DMAYSFAF
HIV gp41	RQGYSP
Ig α (human)	ENLYEGL
EGF receptor (human)	SNFYRAL
CTLA-4 (human)	TGVYVKM
Furin (mouse)	LISYKGL
TGN38 (rat)	ASDYQRL
Lysosomal targeting signals	
Lamp-1 (mouse)	HAGYQTI
Lamp-2a (mouse)	HTGYEQF
Lamp-2a (chicken)	NTGYEQF
Lamp-2b (chicken)	RTGYQSV
Lamp-2c (chicken)	YAGYQTL
Macrosialin (mouse)	QSTYQPL
CD63 (mouse)	RSGYEVM
Acid phosphatase (human)	PPGYRHV
CD3- γ (mouse)	EQVYQPL
HLA-DM (human)	HSSYTPL
GMP-17 (mouse)	RAEYETL
GMP-17 (human)	RPGYETL
TRP-2 (mouse)	RKGYAPL
Cystinosin (human)	RPGYDQL

Overall, $\mu 3B$ preferences more closely resembled those of $\mu 3A$ (Table I), as would have been expected from their high degree of sequence identity.

Preferred Combinations of Two or More Amino Acid Residues—The description of preferences summarized in Table I does not reflect the fact that some combinations of amino acid residues within the same sequence were found at a frequency that was severalfold higher than that expected from the observed frequencies of the individual amino acids. For example, the sequence XXGYXXF was selected by $\mu 2$ with a frequency that was 8-fold higher than that calculated from the observed frequencies of glycine at Y-1 and phenylalanine at Y+3. This means that, even though glycine or phenylalanine were not individually favored at those positions (Fig. 2 and Table I), they

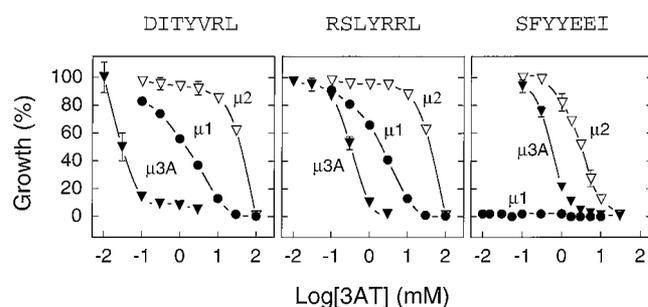


FIG. 3. **3AT growth inhibition assays.** The strength of the interactions between YXX \emptyset sequences and the $\mu 1$, $\mu 2$, and $\mu 3A$ subunits was evaluated by analyzing the effect of increasing concentrations of 3AT on the growth of cotransformed yeast cells. Values on the y axis are the ratios of optical density at 600 nm obtained in the presence or absence of 3AT. Values are the mean \pm SD of triplicate determinations.

were preferred when present in combination. In contrast, the sequence GXXYXRL was never selected by $\mu 2$ even though glycine at Y-3 and arginine at Y+2 were strongly favored individually at their respective positions (Fig. 2 and Table I). Similarly, although $\mu 2$ prefers proline at Y-1 and Y+2, both prolines were never found within the same signal (Fig. 1 and Table I). Thus, cooperative and inhibitory interactions among residues within a signal seem to contribute to the overall binding affinity. The best interpretation of these interactions is that signals must be able to adopt a certain conformation to bind to the medium subunits.

Specificity Overlap—In general, interactions of $\mu 2$ with the signals that it selected were the strongest among all interactions examined, as determined by measurements of β -galactosidase activity (data not shown) and assays for growth in liquid media in the presence of varying concentrations of 3AT (a histidine biosynthesis inhibitor) (14). For example, yeast cells coexpressing full-length $\mu 2$ with the sequences DITYVRL and RSLYRRL were unaffected by concentrations of 3AT of up to 10 mM (Fig. 3), whereas yeast cells coexpressing either $\mu 1$ or $\mu 3A$ with any signal were completely inhibited at that concentration of 3AT (Fig. 3 and data not shown). Most signals selected by $\mu 1$, $\mu 3A$, and $\mu 3B$ bound to some extent to full-length $\mu 2$ (Fig. 1). Even signals that seemed to be specific for $\mu 3A$ (e.g. SFYYEEI) on plate growth assays (Fig. 1) displayed interaction with $\mu 2$ when tested by the more sensitive 3AT resistance assays (Fig. 3). There was also considerable overlap among the signals recognized by $\mu 1$, $\mu 3A$, and $\mu 3B$ (Fig. 1). Taken together, these observations suggest that $\mu 2$ has the highest affinity and broadest specificity for YXX \emptyset signals.

DISCUSSION

The characteristics of YXX \emptyset signals were originally defined from functional analyses of protein sorting in live cells (4-7). These studies established that YXX \emptyset signals are highly degenerate (see Table II), with only the Tyr residue and a bulky hydrophobic residue at the \emptyset position being essential for function. Mutations of residues adjacent to the critical Tyr altered the function of some signals, although the mutated signals retained some level of activity (4-8, 30,31). The identification of the medium subunits of adaptor complexes as recognition molecules for YXX \emptyset signals allowed further characterization of the properties of the signals using protein interaction techniques such as the yeast two-hybrid system and *in vitro* binding assays (10,11). These techniques confirmed at a molecular level that the Tyr and \emptyset residues are critical elements of the signals and that residues adjacent to the Tyr contribute to recognition by the different adaptor medium subunits. However, an important question that was not definitively answered in previous analyses of signal-adaptor interactions was

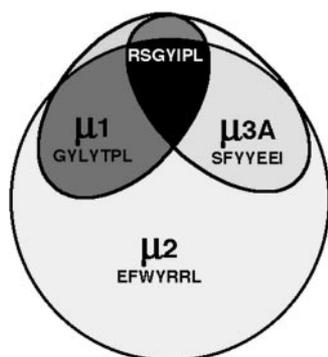


FIG. 4. Conceptual representation of the overlapping specificities of the different medium subunits. Examples of YXXØ sequences in each set are shown. The μ_2 set corresponds to signals binding to the full-length protein encoded by the μ_2^{+E5} cDNA (26).

whether each medium subunit recognizes a distinct set of YXXØ signals. The results of the combinatorial analyses reported here show that each medium subunit does exhibit a preference for certain YXXØ signals; however, there is also considerable specificity overlap among them (Fig. 4).

A preference common to all the medium subunits was proline at position Y+2. Naturally occurring signals containing proline at Y+2 are found in HIV-1 gp41, macrosialin, CD3- γ , HLA-DM, and TRP-2 (Table II). Proline is more conformationally restricted than other amino acid residues due to the fact that the bond between the α -carbon and the nitrogen atom involved in peptide bond formation is fixed within an imidazole ring. Thus, the preference for proline at Y+2 suggests that the signals need to adopt a certain conformation (e.g. a bend in the polypeptide chain) to bind to the medium subunits; this conformation could be a tight turn as previously proposed (32). Another preference common to μ_1 , μ_3A , and μ_3B , but not μ_2 , was arginine at position Y-3. This feature is present in signals from CD1b, lamp-2b, CD63, GMP-17, TRP-2, and cystinosin (Table II), most of which are involved in targeting to late endosomal or lysosomal compartments.

The presence of hydrophobic residues at the X positions was disfavored by all of the medium subunits. This observation is in agreement with the low frequency of hydrophobic residues at those positions in most naturally occurring YXXØ signals (Table II). The preference for polar or charged residues could indicate a participation of those residues in interactions with like groups on the surface of the adaptor medium subunits, where they may be exposed to the aqueous environment.

In addition to these general preferences, each medium subunit displayed a characteristic set of preferences for residues at the X and Ø positions. μ_1 preferences were relatively nondescript, the most salient being leucine or aspartic acid at Y-1. The only signal known to us that fulfills this preference is the sequence ASDYQRL from TGN38 (Table II). Consistent with this, replacement of alanine for aspartic acid in the ASDYQRL sequence decreased interaction with μ_1 (8). Although the AP-1 complex is thought to be involved in lysosomal sorting to the endosomal-lysosomal system from the TGN, to date the only protein for which there is evidence for YXXØ-dependent, AP-1-mediated transport to endosomes or lysosomes is lamp-1 (33). The lamp-1 signal, HAGYQTI, binds μ_1 weakly in yeast two-hybrid assays (8), consistent with the fact that it does not match the μ_1 preferences described here. The intrinsic weakness of this signal is likely compensated by its favorable position within the lamp-1 tail, which makes it an efficient lysosomal targeting signal (33). Given the paucity of data about the effect of mutations on YXXØ- μ_1 interactions and on AP-1-mediated protein sorting events, the significance of the μ_1

preferences described in this study remains unclear.

In contrast to μ_1 , μ_2 preferences could be readily correlated with functional data. As previously demonstrated (11), μ_2 preferred arginine at Y+2, a feature that is found in the internalization signals of at least two proteins, the transferrin receptor and TGN38 (Table II). Also of note is the fact that several other internalization signals (but only few lysosomal targeting signals) have basic residues (arginine or lysine) between the Tyr and Ø residues (Table II). The importance of basic residues at those positions for internalization was previously demonstrated by Kornfeld and co-workers (4, 6), who showed that mutation to alanine of the second lysine residue within the signal YKYSKV of the cation-independent mannose 6-phosphate receptor decreased activity, whereas mutation to arginine increased activity. Mutation of the arginine residue within the TGN38 signal, ASDYQRL, to aspartic acid decreased both interaction with μ_2 and internalization activity (8, 31). Another preference of μ_2 was phenylalanine at Y-2. We did not find any naturally occurring signals that have phenylalanine at that position. However, the YKYSKV sequence from the cation-independent mannose 6-phosphate receptor has a Tyr residue at Y-2 that contributes to the internalization activity of the signal (4, 6). Finally, μ_2 favored leucine at the Ø position, a preference that is in accordance with the work of Gough and Fambrough (34), who showed that signals having leucine at the Ø position were more active in internalization than those that had phenylalanine or valine. Placement of valine for leucine at the Ø position of the TGN38 signal, ASDYQRL, also decreased interaction with μ_2 (8). All of these correlations are consistent with the idea that YXXØ- μ_2 interactions mediate internalization.

Both μ_3A and μ_3B displayed characteristic preferences for acidic residues immediately before and after the critical tyrosine. The preference for glutamic acid at Y+1 is particularly noteworthy because it is a characteristic of proteins targeted to lysosomes or lysosomal-related organelles, such as lamp-2a, CD63, and GMP-17 (Table II). GMP-17 is a multi-spanning membrane protein localized to lytic granules of natural killer cells and cytotoxic T cells (35), a type of secretory organelle with endosomal-lysosomal characteristics. Mouse GMP-17 has glutamic acid at both the Y-1 and Y+1 positions, suggesting that this protein may be particularly suited for interaction with μ_3A or μ_3B . Also interesting is the fact that μ_3A is the only medium chain described to date that displays a preference for glycine immediately before the critical tyrosine residue. As shown in Table II, this is a characteristic of many, although not all, YXXØ signals that function in lysosomal targeting (see also Ref. 30). All of these data, together with recent genetic analyses of AP-3 function (18-22), support a role for μ_3A /AP-3 in protein sorting to lysosomes or specialized endosomal-lysosomal organelles, perhaps by a pathway that is alternative to or partly redundant with that mediated by μ_1 /AP-1.

Although the preferences revealed by the combinatorial analyses provide clues to the roles of the different medium subunits in specific sorting events, the importance of sequence preferences should not be overestimated. Indeed, the medium subunits can bind suboptimal signals in the two-hybrid assays, albeit with lower apparent affinity. These lower affinity interactions may in fact be physiologically relevant, as naturally occurring signals are often not the strongest possible signals *in vivo* (4, 6, 8). Moreover, the sets of signals preferred by each medium chain overlap to a significant extent, especially with μ_2 (Fig. 4). The fact that full-length μ_2 interacts with most YXXØ signals, including those selected by μ_1 , μ_3A , and μ_3B , agrees with the evidence that most naturally occurring signals are capable of mediating internalization (reviewed in Ref. 2).

The broad specificity of recognition by $\mu 2$ may also allow it to serve a quality control function in which proteins that "leak" to the plasma membrane are quickly endocytosed and retargeted to their appropriate intracellular locations. $\mu 1A$, $\mu 3A$, and $\mu 3B$, on the other hand, interact with more limited sets of YXX \emptyset signals, although these sets also intersect (Fig. 4). Thus, these observations argue against the notion that sequence preferences alone determine the specificity of recognition of YXX \emptyset signals within cells. In other words, a YXX \emptyset signal cannot be defined as purely endocytic or lysosomal solely on the basis of its sequence. Rather, a combination of sequence and contextual factors are likely to define the ultimate functional character of the signals.

Among the contextual factors that contribute to the activity of YXX \emptyset signals is the position of the signal relative to the membrane and to the carboxyl terminus of the protein. For example, most lysosomal targeting signals are found at 7–8 residues from the membrane, a position that Rohrer *et al.* (36) have shown to be critical for efficient sorting to lysosomes. In addition, most lysosomal targeting signals occur at the carboxyl terminus of the proteins, a position that enhances interactions with the medium subunits (8). Therefore, the position of YXX \emptyset signals within the tails of lysosomal integral membrane proteins might provide for optimal recognition at sites of sorting to lysosomes (*e.g.* the TGN or endosomes), whereas sequence features would "fine-tune" the specificity of the recognition event. Another factor that is likely to affect signal recognition is the aggregation state of the integral membrane proteins. For instance, microaggregation caused by ligand binding or changes in the pH or ionic environment as the protein traverses different cellular compartments would be expected to increase the avidity of the proteins for membrane-bound adaptors. Finally, newly synthesized proteins encounter the membrane-bound adaptors in sequential fashion as they move through the secretory and endocytic pathways. Because of the localization of AP-1 and AP-3 to TGN and endosomal membranes, these adaptors have the opportunity to interact with signals before proteins are delivered to the plasma membrane.

Acknowledgments—We thank Marie-Christine Fournier and Machie Sakuma for excellent technical assistance and Chean Eng Ooi, Esteban Dell'Angelica, and Chris Mullins for critical review of the manuscript.

REFERENCES

1. Mellman, I. (1996) *Annu. Rev. Cell Dev. Biol.* **12**, 575–625
2. Marks, M. S., Ohno, H., Kirchhausen, T., and Bonifacino, J. S. (1997) *Trends Cell Biol.* **7**, 124–128
3. Schmid, S. L. (1997) *Annu. Rev. Biochem.* **66**, 511–548
4. Canfield, W. M., Johnson, K. F., Ye, R. D., Gregory, W., and Kornfeld, S. (1991) *J. Biol. Chem.* **266**, 5682–5688
5. Collawn, J. F., Kuhn, L. A., Liu, L.-F. S., Tainer, J. A., and Trowbridge, I. S. (1991) *EMBO J.* **10**, 3247–3253
6. Jadot, M., Canfield, W. M., Gregory, W., and Kornfeld, S. (1992) *J. Biol. Chem.* **267**, 11069–11077
7. Naim, H. Y., and Roth, M. G. (1994) *J. Biol. Chem.* **269**, 3928–3933
8. Ohno, H., Fournier, M. C., Poy, G., and Bonifacino, J. S. (1996) *J. Biol. Chem.* **271**, 29009–29015
9. Shiratori, T., Miyatake, S., Ohno, H., Nakaseko, C., Isono, K., Bonifacino, J. S., and Saito, T. (1997) *Immunity* **6**, 583–589
10. Ohno, H., Stewart, J., Fournier, M. C., Bosshart, H., Rhee, I., Miyatake, S., Saito, T., Gallusser, A., Kirchhausen, T., and Bonifacino, J. S. (1995) *Science* **269**, 1872–1875
11. Boll, W., Ohno, H., Songyang, Z., Rapoport, I., Cantley, L. C., Bonifacino, J. S., and Kirchhausen, T. (1996) *EMBO J.* **15**, 5789–5795
12. Dell'Angelica, E. C., Ohno, H., Ooi, C. E., Rabinovich, E., Roche, K. W., and Bonifacino, J. S. (1997) *EMBO J.* **16**, 917–928
13. Chuang, E., Alegre, M. L., Duckett, C. S., Noel, P. J., Vander Heiden, M. G., and Thompson, C. B. (1997) *J. Immunol.* **159**, 144–151
14. Aguilar, R. C., Ohno, H., Roche, K. W., and Bonifacino, J. S. (1997) *J. Biol. Chem.* **272**, 27160–27166
15. Rapoport, I., Miyazaki, M., Boll, W., Duckworth, B., Cantley, L. C., Shoelson, S., and Kirchhausen, T. (1997) *EMBO J.* **16**, 2240–2250
16. Stephens, D. J., Crump, C. M., Clarke, A. R., and Banting, G. (1997) *J. Biol. Chem.* **272**, 14104–14109
17. Zhang, Y., and Allison, J. P. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 9273–9278
18. Ooi, C. E., Moreira, J. E., Dell'Angelica, E. C., Poy, G., Wassarman, D. A., and Bonifacino, J. S. (1997) *EMBO J.* **16**, 4508–4518
19. Simpson, F., Peden, A. A., Christopoulou, L., and Robinson, M. S. (1997) *J. Cell Biol.* **137**, 835–845
20. Cowles, C. R., Odorizzi, G., Payne, G. S., and Emr, S. D. (1997) *Cell* **91**, 109–118
21. Stepp, J. D., Huang, K., and Lemmon, S. K. (1997) *J. Cell Biol.* **139**, 1761–1774
22. Seymour, A. B., Feng, L., Novak, E. K., Robinson, M. S., Swank, R. T., and Gorin, M. B. (1997) *Mol. Biol. Cell* **8**, 227a (abstr.)
23. Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., Neel, B. G., Birge, R. B., Fajardo, J. E., Chou, M. M., Hanafusa, H., Schaffhausen, B., and Cantley, L. C. (1993) *Cell* **72**, 767–778
24. Songyang, S., Margolis, B., Chaudhuri, M., Shoelson, S. E., and Cantley, L. C. (1995) *J. Biol. Chem.* **270**, 14863–14866
25. Luzio, J. P., Brake, B., Banting, G., Howell, K. E., Braghetta, P., and Stanley, K. K. (1990) *Biochem. J.* **270**, 97–102
26. Ohno, H., Poy, G., and Bonifacino, J. S. (1998) *Gene (Amst.)* **2**, 187–193
27. Motulsky, H. (1995) *Intuitive Biostatistics*. Oxford University Press, New York
28. Chen, J.-J., Goldstein, J. L., and Brown, M. S. (1990) *J. Biol. Chem.* **265**, 3116–3123
29. Pevsner, J., Volkandt, W., Wong, B. R., and Scheller, R. H. (1994) *Gene (Amst.)* **146**, 279–283
30. Harter, C., and Mellman, I. (1992) *J. Cell Biol.* **117**, 311–325
31. Humphrey, J. S., Peters, P. J., Yuan, L. C., and Bonifacino, J. S. (1993) *J. Cell Biol.* **120**, 1123–1135
32. Collawn, J. F., Stangel, M., Kuhn, L. A., Esekogwu, V., Jing, S. Q., Trowbridge, I. S., and Tainer, J. A. (1990) *Cell* **63**, 1061–1072
33. Hönig, S., Griffith, J., Geuze, H. J., and Hunziker, W. (1996) *EMBO J.* **15**, 5230–5239
34. Gough, N. R., and Fambrough, D. M. (1997) *J. Cell Biol.* **137**, 1161–1169
35. Medley, Q. G., Kedersha, N., O'Brien, S., Tian, Q., Schlossman, S. F., Streuli, M., and Anderson, P. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 685–689
36. Rohrer, J., Schweizer, A., Russell, D., and Kornfeld, S. (1996) *J. Cell Biol.* **132**, 565–576