Protein Science (1994), *3*:1197-1205. Cambridge University Press. Printed in the USA.

A survey of furin substrate specificity using substrate phage display

DAVID J. MATTHEWS,[1,3] LAURIE J. GOODMAN,[2] CORNELIA M. GORMAN,[2,4] AND JAMES A. WELLS[1]

[1] Department of Protein Engineering and [2] Department of Cell Genetics, Genentech, Inc., South San Francisco, California 94080

(RECEIVED April 8, 1994; ACCEPTED June 3, 1994)

Abstract

The substrate specificity of furin, a mammalian enzyme involved in the cleavage of many constitutively expressed protein precursors, was studied using substrate phage display. In this method, a multitude of substrate sequences are displayed as fusion proteins on filamentous phage particles and ones that are cleaved can be purified by affinity chromatography. The cleaved phage are propagated and submitted to additional rounds of protease selection to further enrich for good substrates. DNA sequencing of the cleaved phage is used to identify the substrate sequence. After 6 rounds of sorting a substrate phage library comprising 5 randomized amino acids (xxxxx), virtually all clones had an RxxR motif and many had Lys, Arg, or Pro before the second Arg. Nine of the selected sequences were assayed using a substrate–alkaline phosphatase fusion protein system. All were cleaved after the RxxR, and some substrates with Pro or Thr in P2 were also found to be cleaved as efficiently as RxKR or RxRR. To further elaborate surrounding determinants, we constructed 2 secondary libraries (xxRx(K/R)Rx and xxRxPRx). Although no consensus developed for the latter library, many of the sequences in the the former library had the 7-residue motif (L/P)RRF(K/R)RP, suggesting that the furin recognition sequence may extend over more than 4 residues. These studies further clarify the substrate specificity of furin and suggest the substrate phage method may be useful for identifying consensus substrate motifs in other protein processing enzymes.

Keywords: furin; substrate phage display; substrate specificity; subtilisin-like protease

Site-specific proteolysis is one of the most widespread and important posttranslational modifications. Many proteins, including neuropeptides, peptide hormones, enzymes, and viral glycoproteins, are synthesized as inactive precursors that require specific endoproteolytic cleavage for maturation. Recently, a novel family of mammalian subtilisin-like proteases has been discovered that appears to be involved in intracellular processing of specific protein precursors (for recent reviews, see Steiner et al., 1992, and Smeekens, 1993).

It is useful to establish which precursor is processed by which enzyme to better understand the regulation of processing in vivo and to design potential therapeutics for affecting a particular processing step. Linking a precursor with a processing enzyme is often done by showing that they are expressed in the same tissues, that they co-localize at the same time and place within the cell, and that the precursor has a sequence that can be cut by the enzyme. This latter aspect is facilitated by understanding the substrate specificity of the enzyme, which can be very time consuming because it traditionally involves the synthesis and kinetic analysis of many peptides or modified protein precursors.

Recently, a technique called substrate phage display has been developed (Matthews & Wells, 1993), which allows one to rapidly survey the specificity of a protease using an iterative selection method (Fig. 1). A randomized substrate linker is placed between a binding domain and the gene III coat protein of the filamentous phage, M13, which displays the fusion protein and packages the gene coding for it inside. The pool of phage particles is bound to an affinity matrix via the binding domain, which in our case is a high-affinity variant of hGH that can bind

Reprint requests to: James A. Wells, Department of Protein Engineering, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080; e-mail: jaw@gene.com.

[3] Present address: Arris Pharmaceutical Corporation, 385 Oyster Point Boulevard, South San Francisco, California 94080.

[4] Present address: MEGABIOS, Two Embarcadero Center, Suite 410, San Francisco, California 94111.

Abbreviations: hGH, human growth hormone; hGHbp, human growth hormone binding protein; AP, *Escherichia coli* alkaline phosphatase; MOI, multiplicity of infection; Boc, *t*-butyloxycarbonyl; MCA, 4-methyl-coumaryl-7-amide; DEAE, diethylaminoethyl; PEG, polyethylene glycol; BSA, bovine serum albumin; MES, 2-(*N*-morpholino) ethanesulfonic acid; Tris, Tris(hydroxymethyl)aminomethane; DMEM, Dulbecco's modified Eagle's medium; AMC, aminomethylcoumarin; PVDF, polyvinylidene difluoride; amino acids are referred to by standard 1- or 3-letter codes.

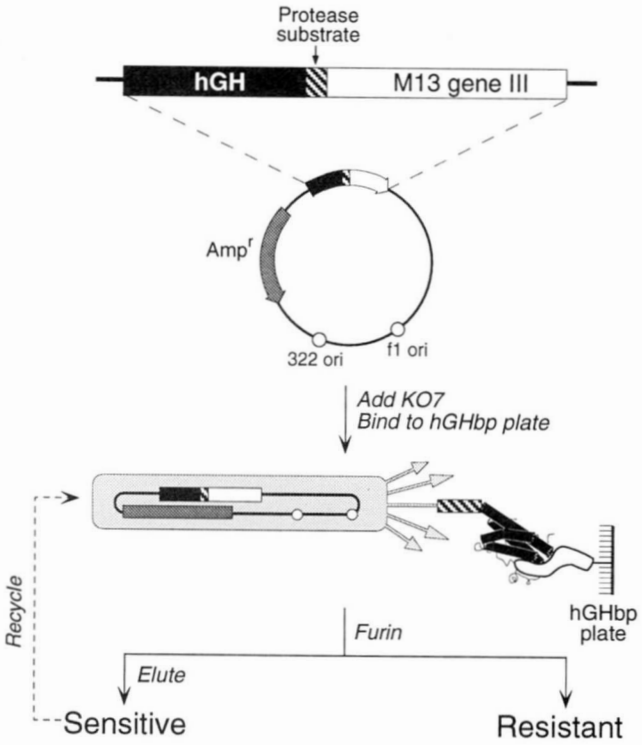


Fig. 1. Scheme for protease substrate phage selection. Each substrate sequence is displayed as a single copy on the surface of phagemid particles, between a truncated form of the gene III protein of M13 and a variant of hGH that binds tightly to hGHbp (see Materials and methods). The substrate phages are bound to hGHbp immobilized on polystyrene plates and cleaved by incubation with furin. The phage containing good substrates within the randomized linker are cleaved, whereas the noncleaved phage remain bound to the immobilized support. By propagating the protease-sensitive phages and repeating the procedure, phage bearing good substrate sequences can be selectively enriched.

to its receptor. The protease of interest is added and the cleaved phage are collected and propagated. To further enrich for good substrates, the cleaved phage are subjected to additional rounds of selection.

We wished to determine if this method could be applied to the family of prohormone processing enzymes, and chose furin as a model because the gene was cloned (Van den Ouweland et al., 1990; Van de Ven et al., 1990) and much was already known about its distribution and specificity. Furin is expressed fairly ubiquitously (Hatsuzawa et al., 1990) and is localized to the trans-Golgi network by sequences in its cytosolic tail (Molloy et al., 1994). In contrast to other prohormone convertases, furin is thought to act on protein precursors that are transported via the constitutive pathway of secretion as opposed to regulated secretory granules. Many constitutively secreted precursor proteins are cleaved with basic residues in the P1, P2, and P4 positions[1] (Hosaka et al., 1991), which matches studies showing that furin requires basic residues at P4 as well as P1 (Hatsuzawa et al., 1992; Molloy et al., 1992). This information has recently been used to design furin inhibitors by synthesis of peptidyl chloroalkylketones containing the RxxR motif (Stieneke-Grober et al., 1992) or by mutating the reactive site of turkey ovomucoid third domain (Lu et al., 1993) or α1-antitrypsin (Anderson et al., 1993).

Kinetic studies with several synthetic peptide-MCA substrates containing Arg at P1 and P4 have suggested that furin has a relatively low k_{cat} (~0.03–2.0 s^{-1}) and that K_m varies widely from ~6 μM (pyroGlu-Arg-Thr-Lys-Arg-MCA) to 0.2 mM (pyroGlu-Arg-Thr-Gln-Arg-MCA). Peptides containing basic residues at the P1 and P2 positions but not at P4 are not recognized by furin (Hatsuzawa et al., 1992). Thus, the substrate-binding site for furin extends over at least 4 residues and perhaps more.

In this work, we expressed and purified a soluble recombinant form of murine furin lacking the transmembrane domain and cytoplasmic tail, and investigated its substrate specificity using substrate phage display (Fig. 1). We confirmed that furin strongly prefers substrates with an RxxR motif as well as having other subsite preferences perhaps extending over 7 residues. These studies provide a general survey of furin substrates and suggest specific ones that can be tested in vitro. Moreover, substrate phage display should be helpful in characterizing the specificity of members of this important family and other processing proteases.

Results

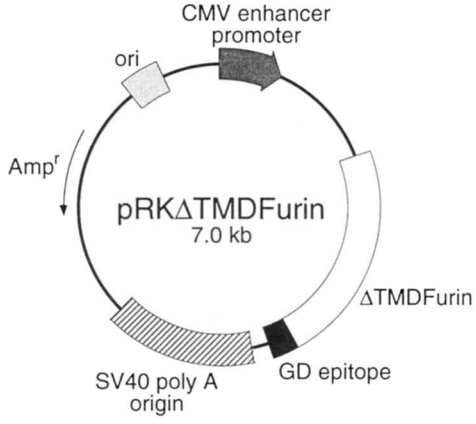
Construction, expression, purification, and kinetic analysis of ΔTMD-furin

The ΔTMD-furin expression plasmid (Fig. 2) was constructed and used for transient transfections of HEK 293 cells. Soluble furin was purified from the conditioned media and assayed as described in the Materials and methods. The presence of furin was verified at each step of the purification by immunoblotting with antibodies directed to the carboxy-terminal GD epitope (Fig. 3). Furin immunoreactivity corresponds to a band migrating at a molecular weight of approximately 90 kDa and could be detected at all steps during the purification (Fig. 3, lanes 2, 3, 4). Supernatants from cells not transfected with the furin expression plasmid did not contain the 90-kDa immunoreactive band (lane 1) and showed no furin activity. The final furin preparation was estimated to be 65–70% pure by SDS-PAGE followed by silver staining. The Michaelis constant (K_m) for hydrolysis of Boc-Arg-Val-Arg-Arg-MCA by our purified furin (67 ± 12 μM) was similar to the value of 57 μM found by Hatsuzawa et al. (1992) for their preparation of soluble murine furin.

Substrate phage selection

Substrate phage selections (Fig. 1) were initially performed with phage containing a known furin substrate sequence (RKKR↓Q: Marriott et al., 1992) that was inserted between the hGH variant and the truncated gene III protein. These phage were bound to hGHbp immobilized on microtiter plates and treated with ~300 nM furin at 37 °C for 2 h. Approximately 4 times as many phage were released when ΔTMD-furin was added compared to a control experiment with buffer alone. This enrichment was less than the 10–100-fold enrichments observed in previous sub-

[1] The nomenclature for substrate residues is NH_2-P*n* . . . P2-P1-P1′-P2′. . . P*n*′-COOH, where cleavage occurs between the P1 and P1′ residues, as described by Schechter and Berger (1967).



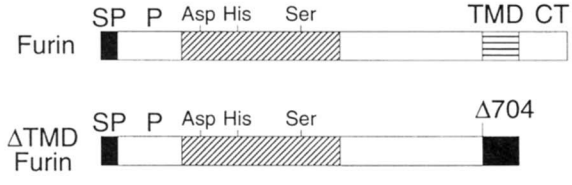


Fig. 2. Schematic representation of the ΔTMD-furin expression plasmid pRKΔTMDFurin. The transmembrane and cytoplasmic domains of furin have been replaced with an antibody epitope comprising the 16 amino-terminal amino acids of herpes simplex virus glycoprotein D. Expression is driven by the human CMV promoter and enhancer, and the furin gene is followed by an SV40 polyadenylation site and origin of replication. The cDNA of the full-length furin (lower diagram) encodes a protein containing a signal sequence (SP), a pro-region (P), and a subtilisin-like domain containing the characteristic Asp, His, and Ser catalytic residues. The wild-type furin is anchored to the Golgi membrane by a transmembrane domain (TMD) followed by a cytoplasmic tail (CT). These later pieces were removed and replaced with the GD epitope tag to allow for soluble expression and immunological detection of the protein.

strate phage studies on 2 other proteases (Matthews & Wells, 1993). Therefore, we performed selection experiments with furin on phage harboring no substrate linker or containing a substrate for the homologous protease PC1 (WSKR↓S), which is not cleaved by furin (Marriott et al., 1992). In either case, no more phage were released on incubation with furin than with buffer alone. Thus, we proceeded to screen a library consisting of 5 randomized codons between di-alanine linkers (AA-xxxxx-AA) for peptides cleaved by ΔTMD-furin.

After sorting the AA-xxxxx-AA library for 6 rounds, 30 clones were isolated and sequenced from the furin-selected substrate library (Table 1). These showed a very high incidence of sequences containing 2 Arg residues separated by 2 amino acids (RxxR). Most of these sequences have Arg in the second and fifth positions within the substrate cassette. These data suggest that Arg is critical at the P1 and P4 positions of the substrate, and that the P1 and P4 residues (rather than the P2 position)

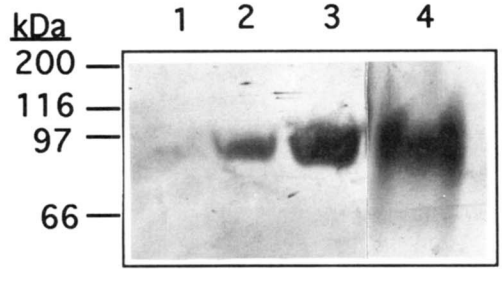


Fig. 3. Immunoblot of ΔTMD-furin purification from 293 cell supernatants. Proteins from conditioned media from untransfected cells (lane 1), from 293 cells transfected with pRKΔTMDFurin (lane 2), from the ammonium sulfate fractionation step (lane 3), and from the DEAE ion-exchange purification step (lane 4) were fractionated on a 10% SDS–polyacrylamide gel, blotted onto PVDF membranes, and detected with anti-5B6 epitope antibody as described in the Materials and methods.

Table 1. *Sequences sensitive to cleavage by furin (purified from transfected cells + furin) or by proteins from cell supernatants not transfected with furin (− furin) that were isolated from AA-xxxxx-AA library after 6 rounds of selection*[a]

+ Furin	− Furin
Sequences fitting **R**xx**R** motif	
P **R** P K **R**	R T N **R** E
I **R** R P **R**	
P **R** T R **R**	
P **R** P R **R**	
I **R** K T **R** (3)	
Y **R** L P **R**	
T **R** K R **R** (2)	
K **R** L P **R**	
K **R** T P **R**	
R **R** T P **R**	
G **R** Q P **R**	
G **R** I K **R**	
P **R** L R **R**	
K **R** T S **R**	
S **R** H I **R**	
V **R** V A **R**	
K **R** T T **R**	
T **R** I K **R**	
R K K **R** D	
R N R **R** S	
R I K **R** W	
Sequences not fitting **R**xx**R** motif	
K G N L E	A N T P P
T S P P L	Q T L P N
T V R S R	D L Q E S
K L E N R	P L R S Q
I G D D L	I S P L S
C T F T T	S K P E R
	G H I E T
	T W T K P
	I M I E P
	M E I M R
	D T N F S
	T L Q P A
	S P L D P
	N H G M P

[a] Numbers in parentheses indicate that a particular DNA sequence was observed more than once.

are the most important specificity determinants for furin. Six clones do not conform to the RxxR motif (bottom of Table 1). These probably represent phage that eluted nonspecifically due to the low level of enrichment of furin substrates per round.

As a control to test the specificity of the selection, we harvested conditioned medium from untransfected 293 cells and subjected it to the same purification procedure as for furin. This preparation was used to conduct a "mock selection" on the substrate phage library, under identical conditions to those used for the furin selection. After 6 rounds of mock selection, 15 sequences were determined. Only 1 sequence was isolated with an RxxR motif and the others showed no similarity to those isolated from the furin selection (Table 1). Proline was slightly overrepresented in the mock selection, but overall a random distribution of sequences was observed.

AP fusion protein assay

In order to determine actual sites of cleavage and to compare relative rates of hydrolysis for some of the isolated RxxR sequences, we prepared hGH–substrate–AP fusion proteins (Fig. 4). Nine different sequences were fused between hGH and AP, bound to the immobilized hGHbp, and treated with furin for various times. N-terminal protein sequencing of the cleaved AP product showed that all substrates containing RxxR were hydrolyzed after the RxxR motif (data not shown). However, the relative rates of hydrolysis varied widely for these sequences, as shown in Figure 5. The sequence TRKRR was the most rapidly cleaved, followed by KRTPR and KRTTR. Several other sequences containing the canonical RxK/RR furin substrate motif were found to be cleaved much less efficiently. The total amount of AP released appears to reach a maximum after about 1 h for many of the substrates (Fig. 5); however, the substrate KRTPR does not follow this trend. In fact, the total amount of AP released increases up to 3 h after initiation of this reaction (data not shown). The release of AP upon furin treatment for all these substrates was much greater than for the release of noncleaved AP as determined from a nonsubstrate sequence, IGDDL (Fig. 5). Thus, the release of AP depends on cleavage of the substrate linker and not cleavage elsewhere in the fusion protein or dissociation of hGH from the hGHbp.

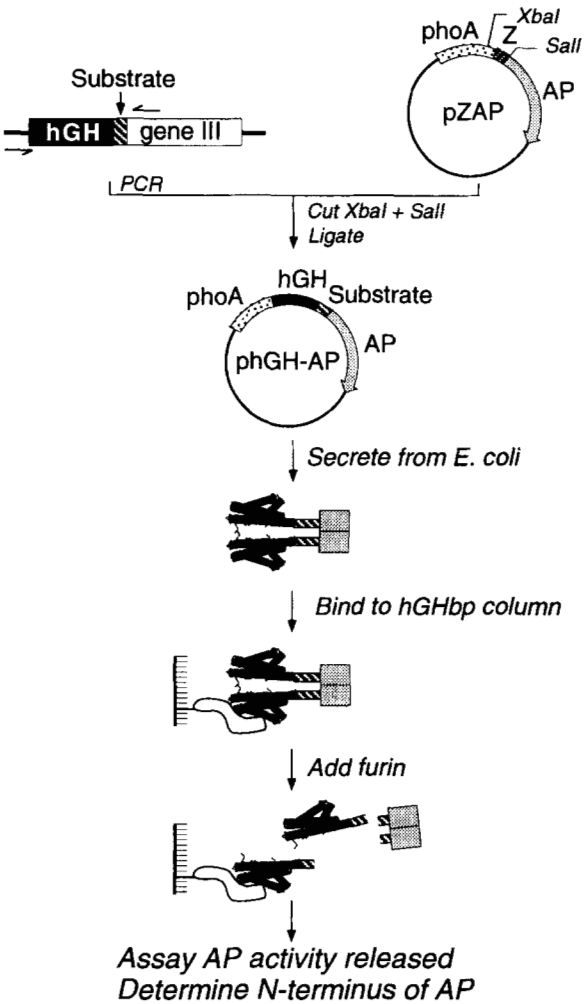


Fig. 4. hGH–AP fusion protein assay used to quantitate the rates and determine the sites of cleavage of individual substrate sequences. Each hGH–substrate sequence is taken out of substrate phage vector by PCR and placed in front of the gene coding for AP, a dimeric enzyme. Plasmids are expressed in *E. coli* and each fusion protein is bound to immobilized hGHbp. Furin is added and allowed to cleave the fusion protein. Supernatants are collected at various times to determine the amount of AP released. The site of cleavage is determined from the N-terminal protein sequence of the cleaved AP isolated after SDS-PAGE and electroblotting onto PVDF membranes as described in the Materials and methods.

Design of extended substrate phage libraries

We aligned the furin-selected RxxR substrates (Table 1) based on the assumption that the conserved Arg residues were at positions P4 and P1 in the substrate region. From this it was noticed that some residues were selected more often at the P2 position compared to what was expected from a random distribution. For example, Lys, Pro, and Arg were isolated at frequencies that were at least 2 standard deviations above random expectation (Fig. 6).

We used this information to design 2 secondary libraries to probe the specificity determinants flanking the RxxR motif more thoroughly. Library A contained the sequence AA-xxRx(K/R)Rx-AA, where x is a random residue and (K/R) means that all oligonucleotides coded for either Lys or Arg at the fifth position. Library B was similar but had Pro fixed in place of K/R (i.e., AA-xxRxPRx-AA). Libraries A and B (containing 3×10^6 and 5×10^5 independent transformants, respectively) were each selected with furin for 5 rounds under identical conditions. Fifteen clones from each library were sequenced (Table 2) and analyzed as above (Fig. 6). From this a consensus emerged for library A, identifying the motif (L/P)RRF(K/R)RP. No obvious consensus developed for Library B.

Discussion and conclusions

Until now, investigations of furin substrate specificity have concentrated on the sequences of natural precursor proteins and a limited set of synthetic peptides derived from these sequences. In this study, we constructed and expressed a truncated form of mouse furin (ΔTMD-furin) and used it to select furin substrates from a vast library of random peptides using the method of substrate phage display. Our principal aim in using this tech-

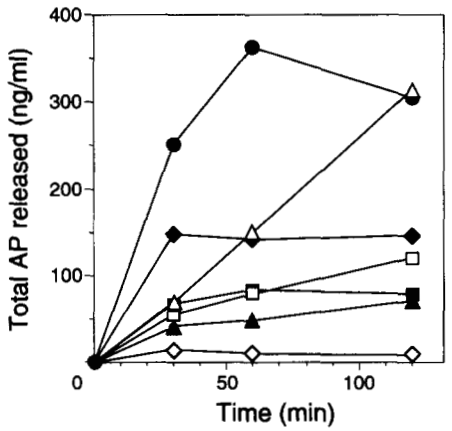


Fig. 5. Results of hGH–AP fusion protein assay. Substrates are as follows: ●, TRKRR; △, KRTPR; ◆, KRTTR; ■, RKKRD; □, PRPRR; ▲, PRPKR; ◇, IGDDL. Data for the substrates PRTRR and IRKTR have been omitted for clarity. These substrates gave results similar to PRPRR and PRPKR.

nique was to provide a broad survey of the substrate specificity of furin as a representative of this important family of processing enzymes. The advantage of the substrate phage approach is that many more sequences can be tested simultaneously compared to conventional methods involving synthetic substrates. Moreover, synthetic substrates usually have a spectroscopic probe at the P1′ position and therefore do not evaluate the specificity for residues that follow the cleaved bond.

There are, however, a number of limitations of the method that one needs to control or be aware of. Firstly, limitations in oligonucleotide synthesis and transformation efficiency mean that one can never be sure that all sequences are represented in the substrate phage library. For example, each of our libraries

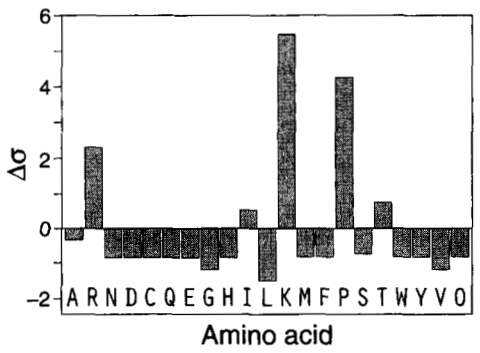


Fig. 6. Statistical distribution of amino acids in the P2 position from the AA-xxxxx-AA library. Assuming a binomial distribution of amino acids, $\Delta\sigma$ shows the difference of the observed frequency from the expected frequency in terms of standard deviations:

$$\Delta\sigma = \frac{Obs(x) - nP(x)}{\{nP(x)[1 - P(x)]\}^{1/2}},$$

where $Obs(x)$ is the number of occurrences of amino acid x in the selected sequences, $P(x)$ is the theoretical probability of amino acid x occurring, and n is the total number of sequences observed. Amino acids are denoted by standard 1-letter codes (O represents the amber stop codon, which is suppressed as glutamine in *E. coli* XL-1).

Table 2. *Sequences sensitive to furin derived after 5 rounds of selection from (A) xxRx(K/R)Rx library and (B) xxRxPRx library*[a]

A. xxRx(K/R)Rx library		
K in P2:	L E R V K R Y	
	L V R R K R L	
	L P R G K R R	
	P R R F K R P	
R in P2:	L V R S R R W	
	S R R F R R P	
	D R R F R R P	
	W F R P R R S	
	P R R P R R G	
	I G R P R R S	
	S R R Y R R P	
	# L R R F R R P	(4)
	# L R R F R R P	
Consensus:	L R R F K R P	
	P R	
B. xxRxPRx library		
	T Q R I P R D	
	H I R K P R W	
	P R R R P R A	
	L G R S P R Y	
	V R R G P R S	
	V S R Y P R S	
	R H R S P R H	
	D D R S P R L	
	P V R A P R S	
	K S R R P R E	
	R R R P P R A	
	L R R I P R H	
	F S R V P R H	
	S A R S P R C	
	K T R I P R G	
	S R R S P R S	
No consensus		

[a] Numbers in parentheses indicate that a particular DNA sequence was observed more than once. Fixed sites are shown in bold; sequences containing more than 1 RxxR motif are underlined. Clones marked # have identical protein sequence but are coded for by different DNA sequences.

contained 4 or 5 NN(G/C) random codons, giving a diversity of 1.1×10^6 to 3.4×10^7 DNA sequences. The AA-xxxxx-AA library had 2×10^6 independent transformants and thus could not have sampled all possible DNA sequences. This may explain why we only found a strong consensus for Arg at the P1 and P4 positions and a weaker consensus for Lys, Arg, or Pro at the putative P2 position. Nonetheless, from this partial consensus we designed secondary libraries (AA-xxRx(K/R)Rx-AA and AA-xxRxPRx-AA) where strongly conserved residues were fixed and other flanking residues were mutated. In these libraries, too, the number of transformants was significantly less than the theoretical number of possible DNA sequences. Additionally, there are proteases in *Escherichia coli* that may cleave some of the substrate phage, thus excluding them from the library. Proline is generally a poor protease substrate when it appears at the P1 or P1′ positions, and it was found about 2.5 times more frequently than expected in the substrate linker when sorting was done in the absence of furin (Table 1). Cysteine residues are often un-

derrepresented in phage libraries (Lowman & Wells, 1993), as was seen here too. This is presumably because unpaired cysteines can cause problems with disulfide bonding in the rest of the fusion protein. Also, some sequences may be more highly expressed, so more copies of these sequences will be present in the substrate phage pool. Finally, there is always a nonspecific background of phage that simply leak through the selections. In fact, we saw a number of nonconsensus sequences even after 6 rounds of selection (Table 1). This can be mitigated by doing even more rounds of selection, especially when the enrichment for substrate over nonsubstrate phage per round is low. Thus, although we have tested many sequences as potential furin substrates, we have certainly not tested all.

From sorting the AA-xxxxx-AA library, a strong consensus developed (the RxxR motif), which was consistent with previous studies showing the enzyme requires Arg at P4 and P1 (Hatsuzawa et al., 1992; Molloy et al., 1992). These same research groups showed that the P2 position exhibits a strong preference for Lys or Arg, and we find this plus a preference for Pro. In fact, a potent inhibitor of furin has recently been constructed with an RxPR motif at the active site (Anderson et al., 1993).

Nine of the substrates isolated from the substrate phage selection were further analyzed using the AP release assay (Fig. 5). The release of AP on incubation with furin was much greater for sequences containing the RxxR motif than for the control sequence IGDDL. The IGDDL sequence was isolated during furin selection, and the fact that it is not cleaved suggests that it (and probably other non-RxxR sequences isolated by selection with furin) were nonspecifically eluted. Some substrates with P2 Pro or Thr are cleaved as efficiently as substrates with P2 Lys or Arg. This suggests that Lys or Arg at the P2 position is less important than Arg at the P1 and P4 positions in determining specificity. However, the best substrates found with Thr or Pro in the P2 position also contained Lys in P5 (Fig. 5), suggesting the possibility of an additional subsite in the enzyme that recognizes basic residues. Interestingly, the furin autocatalytic site (KRRAKRD) also contains a basic residue at the P5 position (Leduc et al., 1992). In view of the different steric and chemical properties of Lys/Arg compared to Pro, the P2 Lys/Arg and P2 Pro substrates may exhibit distinctly different binding to the furin active site and different enzyme kinetics. The unusual behavior of the KRTPR substrate sequence in the AP fusion assay provides some preliminary evidence suggesting this may be the case. As noted above, the total amount of AP released in the fusion protein assay does not appear to saturate for the KRTPR substrate. One possible explanation for this is that product inhibition may be much less for KRTPR substrate than for the other substrates tested. Detailed kinetic analysis with synthetic peptide substrates should help in understanding these results.

Having established the RxxR motif, we evaluated the flanking residues in secondary libraries. When we sorted the secondary library with P2 Pro (AA-xxRxPRx-AA), no further consensus was obtained after 5 rounds of selection. However, the AP assay data described above suggest that some P2 Pro substrates can be cleaved as efficiently as P2 Lys/Arg substrates; therefore, P2 Pro substrates may be able to tolerate more widespread substitutions in the P6, P5, P3, and P1′ positions than P2 Lys/Arg substrates.

On sorting the secondary library with P2 fixed as Lys or Arg (AA-xxRx(K/R)Rx-AA), we obtained the consensus sequence (L/P)RRF(K/R)RP. Ten of the 16 clones contained the motif xRRx(K/R)RP and 8 of these were xRRF(K/R)RP. Although Arg at the fifth position in the substrate cassette was isolated more frequently than Lys, the difference was not statistically significant. Arg in the second position was always co-selected with Pro in the seventh position. Because no subtilisins are known that cleave Pro at the P1′ position, it is likely that these substrates are hydrolyzed as (L/P)RRF(K/R)↓RP. Nine of 10 substrates containing the xRRx(K/R)RP motif had Arg rather than Lys at position 5 in the cassette, and it is known that P1 Arg substrates are better than P1 Lys substrates. We attempted to isolate an AP fusion protein containing the substrate LRRFRRP. However, it could not be obtained in sufficient quantities for assay because of degradation in the *E. coli* periplasm.

Co-selection was also observed at the fourth and seventh positions within the substrate cassette. When Phe or Tyr were observed at position 4, Pro was invariably found at position 7. As described above, we believe that Phe/Tyr is likely to occupy the P2 position in these sequences with Pro at P2′. This combination could not have been observed in the initial library because the P2′ position was always in the di-alanine linker region. However, when position 4 was Pro, the corresponding residue at position 7 was Gly or Ser. In this case, it is possible that Pro occupies the P3 position with Gly/Ser at P1′. Pro was also observed at P3 in the initial library, in which case the P1′ residue was also a small amino acid (Ala, in the di-alanine linker region).

Determining the precise roles of furin and other intracellular endoproteases is essential to understanding how prohormones and other bioactive peptides are regulated in vivo. Such regulation is likely to be a complex process, in view of the many Kex2-like proprotein convertases and their varied tissue distribution. The specificity of enzyme–substrate interactions is critically important to these studies. In the case of furin, we have demonstrated the utility of the substrate phage display method in isolating rare and, in some cases, hitherto unknown furin substrates from a vast library of peptides.

Materials and methods

Construction of a secreted form of furin (ΔTMD-furin)

The truncated form of furin (ΔTMD-furin), which contains a carboxyl-terminal epitope from the amino-terminal portion of the herpes simplex virus glycoprotein D (Paborsky et al., 1990), was constructed as follows. A vector containing the human cytomegalovirus enhancer and promoter followed by an intron and containing an SV40 polyadenylation signal (Marriott et al., 1992) was digested with *Xba* I and *EcoR* I and a 4.7-kb fragment was isolated. The same vector containing the full-length cDNA sequence of mouse furin was digested with *Xba* I and *Pst* I, and a 2.13-kb fragment was isolated that lacked the last 267 bp, encoding the transmembrane and cytoplasmic domains. PCR (Saiki et al., 1985) was used to amplify a 135-bp fragment from a plasmid containing 16 amino acids of the herpes simplex virus glycoprotein D (Paborsky et al., 1990). The primers used in the amplification reaction created *Pst* I and *EcoR* I restriction enzyme sites for cloning into the expression vector pRK7 (Gorman et al., 1990). The sequences of the forward and reverse primers were: 5′-AAAATGCAACTGCAGGGCATCGAGGG CAGGGGCACAGAGAGCTCTGGTGGTC-3′ and 5′-TTTT GAATTCGTGAACTTGCGTACTAGC-3′, respectively. The

amplified product from the PCR reaction was digested with *Pst* I and *EcoR* I. This fragment and the fragments isolated above were cloned in a 3-part ligation into pRK7 (Gorman et al., 1990) to form the vector pRKΔTMD-furin (Fig. 2).

Expression in HEK 293 cells

HEK 293 cells, an adenovirus-transformed human embryonic kidney cell line (Gorman et al., 1990), were propagated as monolayers on Corning tissue culture plates in DMEM supplemented with 10% fetal bovine serum at 37 °C in a humidified atmosphere of air:CO_2 (95:5). Cells were grown to 80% confluence on 15 × 100-cm tissue culture plates (Falcon). An expression plasmid containing ΔTMD-furin described above was transiently transfected into HEK 293 cells using the calcium phosphate precipitation method previously described (Gorman et al., 1990). Twelve hours after transfection, the cells were washed with PBS and serum-free medium was added. Conditioned medium was harvested at 48 h posttransfection.

Purification of ΔTMD-furin

ΔTMD-furin was purified using a modification of the protocol by Hatsuzawa et al. (1992). Purification procedures were performed at 4 °C and all buffers contained 1 mM $CaCl_2$. Harvested conditioned medium (375 mL) was brought to 40% saturation with ammonium sulfate, stirred for 1 h, and centrifuged at 10,000 × *g* for 30 min. Ammonium sulfate was then added to the supernatant to 60% saturation. Following centrifugation at 10,000 × *g* for 30 min, the pellet was dissolved in 10 mL of 10 mM MES, pH 6.0, and dialyzed overnight against 10 mM MES, pH 7.0. The dialyzed solution from the ammonium sulfate fractionation was applied to a column (1.0 × 20 cm) of DEAE-Toyopearl 650S (Tosah), which was equilibrated with 10 mM MES/NaOH, pH 7.0, containing 50 mM NaCl. The sample was eluted with a linear 120-mL gradient of 50–500 mM NaCl in the same buffer. Fractions (1.5 mL) were collected at a flow rate of 2 mL/min. On DEAE ion-exchange chromatography, furin activity appeared as 1 peak that eluted at 100–150 mM NaCl. Following purification on DEAE ion-exchange, the specific activity rose from 774 to 3,860 nmol of Boc-Arg-Val-Arg-Arg-MCA hydrolyzed/h/mg of total protein.

Assays

Samples (50 μL) at each purification step were incubated with Boc-Arg-Val-Arg-Arg-MCA (Peninsula Laboratories, Inc.) (20 nmol) in 200 μL of 100 mM MES, pH 7.0, containing 1 mM $CaCl_2$ at 37 °C for 4 h. The reaction was stopped by the addition of 3 mL of 5 mM EDTA. The amount of MCA released from the substrate above was measured in a fluorescence spectrophotometer with an excitation at 380 nm and an emission at 460 nm. The amount of MCA released was quantified by comparing with a standard curve of free AMC (Peninsula). We assumed the k_{cat} value reported by Hatsuzawa et al. (1992) for highly purified furin (k_{cat} = 0.2 s^{-1}) to estimate enzyme concentration.

Enzyme kinetics were measured using a Labsystems Fluoroscan II microplate fluorimeter in conjunction with the Deltasoft II data analysis program (Biometallics, Inc.). Boc-Arg-Val-Arg-Arg-MCA (180 μL) was added to wells of a microtiter plate to give final concentrations ranging from 5 to 200 μM. Fixed concentrations of AMC, ranging from 0.05 to 50 μM, were also loaded in the plate to provide a linear range of fluorescent standards. The plate was equilibrated at 37 °C for 10 min, then 20 μL of enzyme was added to wells containing substrate. Initial rates of AMC release were measured for 40 min and found to be linear. Data were fit to the Michaelis-Menten equation using the program Kaleidagraph (Abelbeck software).

Immunoblot analysis

Protein samples at each stage of the purification scheme were solubilized in SDS-PAGE sample buffer and electrophoresed on precast 10% Tris-glycine SDS gels (Novex). Following electrophoresis, the proteins were electroblotted onto a PVDF membrane, blocked with 2% BSA in TBST buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% Tween 20), then incubated 1 h with a 1:500 dilution of antiserum directed against the herpes simplex virus glycoprotein D epitope. The blot was washed 3 times (5 min per wash) in TBST and incubated 1 h with a 1:1,000 dilution of AP-conjugated goat anti-mouse IgG (Promega), washed again, and developed with nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl-1-phosphate reagent (Promega).

Construction of furin substrate phage, PC1 substrate phage, and substrate phage libraries

DNA sequences coding for known furin and PC1 substrates from human relaxin (Marriott et al., 1992) were introduced between a variant hGH gene and truncated gene III in the phagemid vector pDM0612 (Matthews & Wells, 1993) by cassette mutagenesis. The hGH variant contains the mutations R167N, D171N, K172S, F176Y, I178T, F10A, M14W, H18D, and H21N, and has been shown to bind very tightly to the hGHbp (Lowman et al., 1991). Two pairs of complementary oligonucleotides were used: Pair 1 (see below), directing synthesis of the PC1 substrate (G)PTWSKRSLSQS, and Pair 2 (see below), directing synthesis of the furin substrate (G)PSRKKRSQLYSQS. Both the above cassettes contain *Apa* I and *Sal* I sticky ends; the codon for (G) at the start of each peptide sequence is formed on ligation into pDM0612. The oligonucleotide pairs were annealed by heating equimolar amounts to 97 °C and cooling slowly to room temperature. The cassette was then ligated into the *Apa* I-*Sal* I backbone of pDM0612. The substrate phage library consisting of 5 randomized codons flanked by di-alanine linkers has

5′-CACTTGGAGCAAACGCAGCCTGAGCCAG-3′
3′-CCGGGTGAACCTCGTTTGCGTCGGACTCGGTCAGCT-5′ (Pair 1)

5′-CAGCCGAAAAAAGCGACAGCTCTACAGCCAG-3′
3′-CCGGGTCGGCTTTTTTCGCTGTCGAGATGTCGGTCAGCT-5′ (Pair 2)

been previously described (Matthews & Wells, 1993). This library contained 2×10^6 independent transformants. Secondary substrate libraries were constructed with synthetic oligonucleotides, essentially as described above for the fixed substrate cassettes. The oligonucleotide cassettes used were: Pair 3 (see below) for the xxRxK/RRx library and Pair 4 (see below) for the xxRxPRx library, where N = G + A + T + C and S = G + C.

Substrate phage selection

Substrate phage selection was performed essentially as described by Matthews and Wells (1993) (Fig. 1). Phage particles were produced by growing the transformed *E. coli* in 50 mL 2YT medium overnight in the presence of M13K07 helper phage, which was added at an MOI of ~100. Polystyrene plates (Nunc) containing 6 wells were coated with hGHbp (2 μg/mL in 50 mM carbonate, pH 9.6, overnight at 4 °C) and blocked by incubating with PBS containing 0.5% (w/v) BSA. Phage (approximately 10^{11} colony-forming units [cfu]) were bound to the plates in 1 mL 10 mM Tris, pH 7.6, 100 mM NaCl, 1 mM EDTA, by gently shaking for 2 h at room temperature. The plates were washed extensively over a 1-h period, equilibrated in 100 mM MES, pH 7.0, 5 mM $CaCl_2$, and incubated with ~300 nM furin in the same buffer for 2 h at 37 °C. A control incubation was performed with no enzyme added. Following enzymatic elution, phage were added to 1 mL of a fresh *E. coli* culture and tumbled at 37 °C for 1 h. The infected cells were then centrifuged at 4,000 rpm for 10 min to remove any remaining furin, and then resuspended in 1 mL 2YT medium. The cells were then grown overnight in the presence of helper phage as described above, and the process was repeated up to 6 times.

Construction and assay of AP fusion proteins

Genes coding for hGH and the substrate region were isolated from individual clones using PCR with the primers 5′-TGTCAC GGCCGAGACTTATAGTCGC-3′ and 5′-CACCGCCAGTCG ACCCAGGACCACC-3′ (Fig. 5). The products were isolated and cleaved with *Xba* I and *Sal* I, then gel purified and ligated into the vector pZAPDM (Matthews & Wells, 1993). The resulting constructs were transformed into *E. coli* KS330 cells (Strauch & Beckwith, 1988) and used to express hGH-substrate-AP fusion proteins as previously described (Matthews & Wells, 1993). Following precipitation with 45% (w/v) ammonium sulfate and resuspension of the fusion protein preparations, a second precipitation was performed with 25% (w/v) PEG3350 for 1 h at 4 °C. This was found to efficiently remove a large hGH contaminant that presumably resulted from in vivo cleavage of the fusion protein. After recentrifugation, the final protein pellet was resuspended in 0.5 mL shockate buffer and quantitated by SDS-PAGE in conjunction with laser scanning densitometry. hGHbp was coupled to Affigel 15 (Bio-Rad) using methods recommended by the manufacturer. Equal amounts of each fusion protein (approximately 5 μg as determined by laser scanning densitometry in comparison to an hGH standard) were bound to 50 μL gel slurry at 4 °C for 2 h. The slurry was washed twice in 10 mM Tris, pH 7.6, 1 mM EDTA, and then in furin assay buffer (100 mM MES, pH 7.0, 5 mM $CaCl_2$). Furin (~120 nM in 1 mL) was incubated with the bead slurry, and 100-μL aliquots of supernatant were removed at 30, 60, and 120 min. To assay AP activity, 20 μL of each sample was incubated with 180 μL 6.6 mM *p*-nitrophenylphosphate (pNPP) in wells of a microtiter plate and the release of *p*-nitroaniline (pNA) measured by its absorbance at 405 nm. AP standards were included to provide a linear standard curve. For NH_2-terminal sequence analysis, AP cleavage products were isolated essentially as described above, but 100 μL of gel slurry was used with 100 μL of fusion protein (≥5 μg) and 100 μL of 600 nM furin was used for elution. The resulting AP digestion product was isolated by electroblotting onto PVDF membranes (Matsudaira, 1987), and the NH_2-termini were determined with a protein sequencer (Applied Biosystems 473 or 477).

Acknowledgments

We thank Bill Henzel and Suzy Wong for N-terminal protein sequencing and Wayne Anstine for help with graphics.

References

Anderson ED, Thomas L, Hayflick JS, Thomas G. 1993. Inhibition of HIV-1 gp160-dependent membrane fusion by a furin-directed alpha 1-antitrypsin variant. *J Biol Chem 268*:24887–24891.

Gorman CM, Gies DR, McCray G. 1990. Transient production of proteins using an adenovirus transformed cell line. *DNA Protein Eng Tech 2*:3–10.

Hatsuzawa K, Hosaka M, Nakagawa T, Nagase M, Shoda A, Murakami K, Nakayama K. 1990. Structure and expression of mouse furin, a yeast Kex2-related protease. *J Biol Chem 265*:22075–22078.

Hatsuzawa K, Nagahama M, Takahashi S, Takada K, Murakami K, Nakayama K. 1992. Purification and characterization of furin, a Kex2-like processing endoprotease, produced in Chinese hamster ovary cells. *J Biol Chem 267*:16094–16099.

Hosaka M, Nagahama M, Kim WS, Watanabe T, Hatsuzawa K, Ikemizu J, Murakami K, Nakayama K. 1991. Arg-X-Lys/Arg-Arg motif as a signal for precursor cleavage catalyzed by furin within the secretory pathway. *J Biol Chem 266*:12127–12130.

Leduc R, Molloy SS, Thorne BA, Thomas G. 1992. Activation of human furin precursor processing endoprotease occurs by an intramolecular autoproteolytic cleavage. *J Biol Chem 267*:14304–14308.

Lowman HB, Bass SH, Simpson N, Wells JA. 1991. Selecting high-affinity binding proteins by monovalent phage disiplay. *Biochemistry 30*: 10832–10838.

Lowman HB, Wells JA. 1993. Affinity maturation of human growth hormone by monovalent phage display. *J Mol Biol 234*:564–578.

Lu W, Zhang W, Molloy SS, Thomas G, Ryan K, Chiang Y, Anderson S, Laskowski M Jr. 1993. Arg15-Lys17-Arg18 turkey ovomucoid third domain inhibits human furin. *J Biol Chem 268*:14583–14585.

Marriott D, Gillece-Castro B, Gorman C. 1992. Prohormone convertase-1 will process prorelaxin, a member of the insulin family of hormones. *Mol Endocrinol 6*:1441–1450.

Matsudaira P. 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J Biol Chem 262*: 10035–10038.

Matthews DJ, Wells JA. 1993. Substrate phage: Selection of protease substrates by monovalent phage display. *Science 260*:1113–1117.

5′-CGCTGCTNNSNNSCGTNNSARACGTNNSGCTGCTCCTGGC-3′
3′-CCGGGCGACGANNSNNSGCANNSTYTGCANNSCGACGAGGACCGAGCT-5′ (Pair 3)

5′-CGCTGCTNNSNNSCGTNNSCCGCGTNNSGCTGCTCCTGGC-3′
3′-CCGGGCGACGANNSNNSGCANNSGGCGCANNSCGACGAGGACCGAGCT-5′ (Pair 4)

Molloy SS, Bresnahan PA, Leppla SH, Klimpel KR, Thomas G. 1992. Human furin is a calcium-dependent serine endoproteinase that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. *J Biol Chem 267*:16396–16402.

Molloy SS, Thomas L, VanSlyke JK, Stenberg PE, Thomas G. 1994. Intracellular trafficking and activation of the furin proprotein convertase: Localization to the TGN and recycling from the cell surface. *EMBO J 13*:18–33.

Paborsky LR, Fendly BM, Fisher KL, Lawn RM, Marks BJ, McCray G, Tate KM, Vehar GM, Gorman CM. 1990. Mammalian cell transient expression of tissue factor for the production of antigen. *Protein Eng 3*:547–553.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985. Enzymatic amplification of beta-globin sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science 230*:1350–1354.

Schechter I, Berger A. 1967. On the size of the active site in proteases. *Biochem Biophys Res Commun 127*:157–162.

Smeekens SP. 1993. Processing of protein precursors by a novel family of subtilisin-related mammalian endoproteases. *Bio/Technology 11*:182–186.

Steiner DF, Smeekens SP, Ohagi S, Chan SJ. 1992. The new enzymology of precursor processing endoproteases. *J Biol Chem 267*:23435–23438.

Stieneke-Grober A, Vey M, Angliker H, Shaw E, Thomas G, Roberts C, Klenk HD, Garten W. 1992. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *EMBO J 11*:2407–2414.

Strauch KL, Beckwith J. 1988. An *Escherichia coli* mutation preventing degradation of abnormal periplasmic proteins. *Proc Natl Acad Sci USA 85*:1576–1580.

Van den Ouweland AMW, Van Duijnhoven HLP, Keizer GD, Dorssers LCJ, Van de Ven WJM. 1990. Structural homology between the human *fur* gene product and the subtilisin-like protease encoded by yeast Kex2. *Nucleic Acids Res 18*:644.

Van de Ven WJM, Voorberg J, Fontijn R, Pannekoek H, Van den Ouweland AMW, Van Duijnhoven HLP, Roebroek AJM, Siezen RJ. 1990. Furin is a subtilisin-like proprotein processing enzyme in higher eukaryotes. *Mol Biol Rep 14*:265–275.