Structure 14, 653–660, April 2006 DOI 10.1016/j.str.2006.02.006

Protease Accessibility Laddering: A Proteomic Tool for Probing Protein Structure

Technical Advance

Svetlana Dokudovskaya,[1] Rosemary Williams,[1] Damien Devos,[2] Andrej Sali,[2,4] Brian T. Chait,[3,5] and Michael P. Rout[1,6,*]

[1] Laboratory of Cellular and Structural Biology
The Rockefeller University
1230 York Avenue
New York, New York 10021

[2] Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry and
California Institute for Quantitative Biomedical Research
Byers Hall, 1700 4th Street, Suite 503B
University of California, San Francisco
San Francisco, California 94143

[3] Laboratory of Mass Spectrometry and Gaseous Ion Chemistry
The Rockefeller University
1230 York Avenue
New York, New York 10021

Summary

Limited proteolysis is widely used in biochemical and crystallographic studies to determine domain organization, folding properties, and ligand binding activities of proteins. The method has limitations, however, due to the difficulties in obtaining sufficient amounts of correctly folded proteins and in interpreting the results of the proteolysis. A new limited proteolysis method, named protease accessibility laddering (PAL), avoids these complications. In PAL, tagged proteins are purified on magnetic beads in their natively folded state. While attached to the beads, proteins are probed with proteases. Proteolytic fragments are eluted and detected by immunoblotting with antibodies against the tag (e.g., Protein A, GFP, and 6×His). PAL readily detects domain boundaries and flexible loops within proteins. A combination of PAL and comparative protein structure modeling allows characterization of previously unknown structures (e.g., Sec31, a component of the COPII coated vesicle). PAL's high throughput should greatly facilitate structural genomic and proteomic studies.

Introduction

Limited proteolysis is a powerful and widely used technique to study protein structure, folding, and dynamics (Fontana et al., 2004; Hubbard, 1998). This method is based on the proteolytic susceptibility of specific, sufficiently exposed, and flexible chain segments in a folded protein. Preferred sites of limited proteolysis generally require local unfolding of ten or more residues (Hubbard, 1998).

The original hypothesis, proposed by Neurath, was that limited proteolysis occurs at "hinges and fringes" such as exposed surface loops and domain-linking segments (Neurath, 1980). These hinge regions are usually more flexible than the more compact domains forming the rest of the polypeptide chain. Therefore, limited proteolysis is one of the most suitable techniques to identify and produce individual domains for further structural and functional characterization (see, for example, Cohen and Chait, 2001). Even for proteins with known structures, limited proteolysis can provide important information about folding intermediates (Fontana et al., 2004). In addition, limited proteolysis has been used to study molecular features of protein aggregates associated with severe diseases, such as Alzheimer's disease and Parkinson's disease, and type 2 diabetes (Polverino de Laureto et al., 2003). The presence or absence of a bound ligand can also affect the susceptibility of a protein segment to proteases, resulting in either increased or decreased accessibility. Monitoring of ligand binding by limited proteolysis includes the study of the effects induced by small ligands such as Ca^{2+} ions (Pedigo and Shea, 1995), as well as footprinting of protein-protein (Scaloni et al., 1998; Zhong et al., 1995) and DNA-protein interactions (Cohen et al., 1995).

Despite the utility of limited proteolysis, the method has several technical limitations. First, in order to perform proteolysis, it is necessary to obtain a pure protein in its native folded state, which can be technically difficult, especially for large multidomain proteins. Second, in many cases, the proteins of interest are expressed and purified from bacterial expression systems, which often fail to add posttranslational modifications necessary for normal protein folding. Third, monitoring proteolysis products requires either a sufficiently large amount of product to allow visualization by gel staining, or raising specific antibodies against the amino- or carboxy-terminal part of a protein to be detected by Western blot analysis; these can be time-consuming procedures, especially for novel and rare proteins.

Here we describe a rapid and sensitive limited proteolysis method which avoids the complications mentioned above and can be applied to the study of isolated individual proteins as well as proteins in complexes. In this new variant of the limited proteolysis approach, which we name protease accessibility laddering (PAL), genomically tagged natively folded proteins are isolated from yeast on magnetic beads. Widely used tags, such as PrA, GFP, and 6×His, can be used to attach the protein molecules to the beads. While attached to the magnetic beads, the tagged proteins are probed with proteases. The products of proteolysis are monitored by immunoblotting using antibodies against the tag. We show that PAL allows us to identify proteolytically susceptible regions of the proteins. We also demonstrate how a combination of PAL and comparative protein structure modeling can be applied for the structural characterization of proteins and protein complexes.

*Correspondence: rout@rockefeller.edu
[4] Lab address: http://salilab.org
[5] Lab address: http://prowl.rockefeller.edu
[6] Lab address: http://www.rockefeller.edu/labheads/rout/rout-lab.php

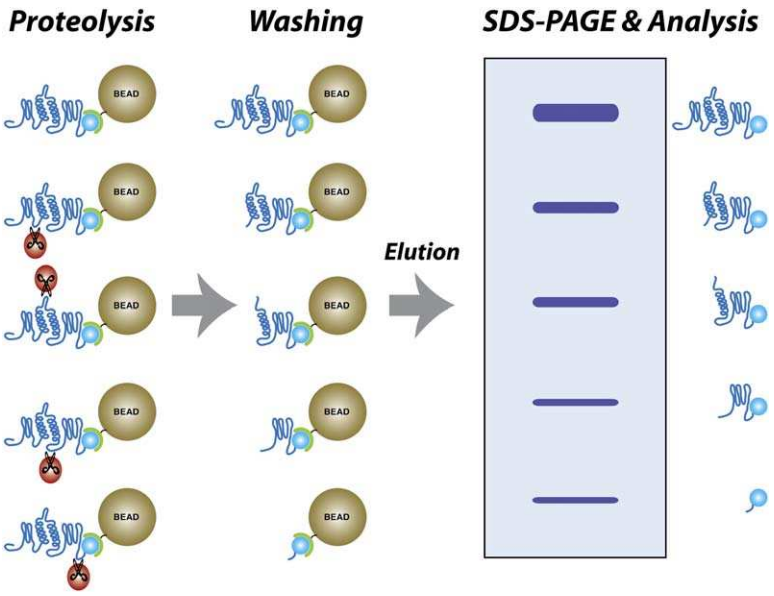

Figure 1. Schematic Representation of the PAL Technique

Proteins attached via their carboxy-terminal tags (blue ball) to an appropriate antibody (in green) conjugated to magnetic beads, are cleaved by a protease (scissors) at their exposed regions. After washing, only the carboxy-terminal fragments remain attached to the beads. These fragments are eluted and can be analyzed by SDS-PAGE and immunoblotting.

Results

Experimental Design

The principle of the PAL method is illustrated in Figure 1. Cell lysates, prepared from yeast strains carrying genomically tagged proteins of interest, are affinity purified on magnetic beads covered with the specific antibodies against the tag. After binding, the beads are recovered and washed to remove excess unbound proteins. Depending on extraction and washing conditions, a protein of interest can be isolated in native form either alone or in a complex with other proteins. The washed beads with attached proteins are treated with the specific proteases Asp-N, Lys-C, or trypsin (other proteases also prove useful in certain cases). After incubation, the proteolytic reactions are stopped at different time points by rapidly harvesting the beads and washing away both the enzyme and the released proteolytic fragments. At this point, only the tagged fragments remain on the beads. Finally, these tagged proteolytic fragments are eluted, separated by SDS-PAGE, and visualized by immunoblotting with labeled rabbit IgG (to detect PrA fragments) or with anti-GFP antibodies. With the carboxy-terminal tagged constructs that we used, the bands on the blot all correspond to carboxy-terminal fragments containing the tag. We verified that the enzyme concentrations and times of reaction described in this paper do not significantly affect the stability of the PrA tag (data not shown). An advantage of using carboxy-terminally tagged proteins is that the resulting tagged proteolytic fragments have free amino termini, which allowed us to use standard amino-terminal Edman sequencing to determine the exact cleavage site for a given proteolytic fragment. The molecular weights of the fragment bands were also determined by comparison with MW standards. By finding the nearest cleavable amino acid in the protein sequence that could generate a fragment of that size, we were able to estimate the cleavage site in those cases where we were unable to obtain an unambiguous amino-terminal sequence result. Although such estimates are susceptible to anomalous gel migration, they nevertheless proved to be within an average of ~7% of the actual fragment molecular weight (by comparison with the Edman data for the same fragments), allowing on average a determination of the cleavage site within ~50 amino acid residues. Depending on the initial results, further details on the accessibility and dynamics of different sites of the protein can be obtained by proteolytic screening with different enzyme:protein ratios, times of reaction, temperatures, or even enzymes. The time required for the PAL process, from growing the initial cultures to readout of the proteolytic sites on a gel, is less than a week.

To investigate the potential of our method, we chose to work with protein components of coated vesicles (clathrin and Sec31), nuclear pore complexes (nucleoporin Nsp1), and nucleocytoplasmic transport factors (Kap95 and Kap60). The choice of proteins was dictated by our interest in the structure and function of the nuclear pore complex and its components, and our recent findings concerning the potential common evolutionary origin of nuclear pore complexes (NPCs) and coated vesicles (Devos et al., 2004).

Validation of the Method

In our initial evaluation of the PAL procedure, we tested two proteins whose domain organization is well defined. The first protein studied was Nsp1. This protein is a member of the phenylalanine-glycine (FG) family of nucleoporins (nups) and contains multiple natively unfolded FG repeats at its amino-terminal end (Denning et al., 2003), followed by well-defined carboxy-terminal coiled-coil regions (Bailer et al., 2001). As expected, Nsp1-PrA was susceptible to proteolysis along its FG region and in the linker regions between the four coiled coils in its carboxyl terminus (Figure 2A).

The second protein that we used to test PAL was clathrin, for which an atomic structure is available (Fotin et al., 2004; ter Haar et al., 1998; Ybe et al., 1999). The clathrin heavy chain (CHC) consists of an amino-terminal seven-blade β propeller, connected via an α-helical zigzag linker to an α solenoid containing multiple

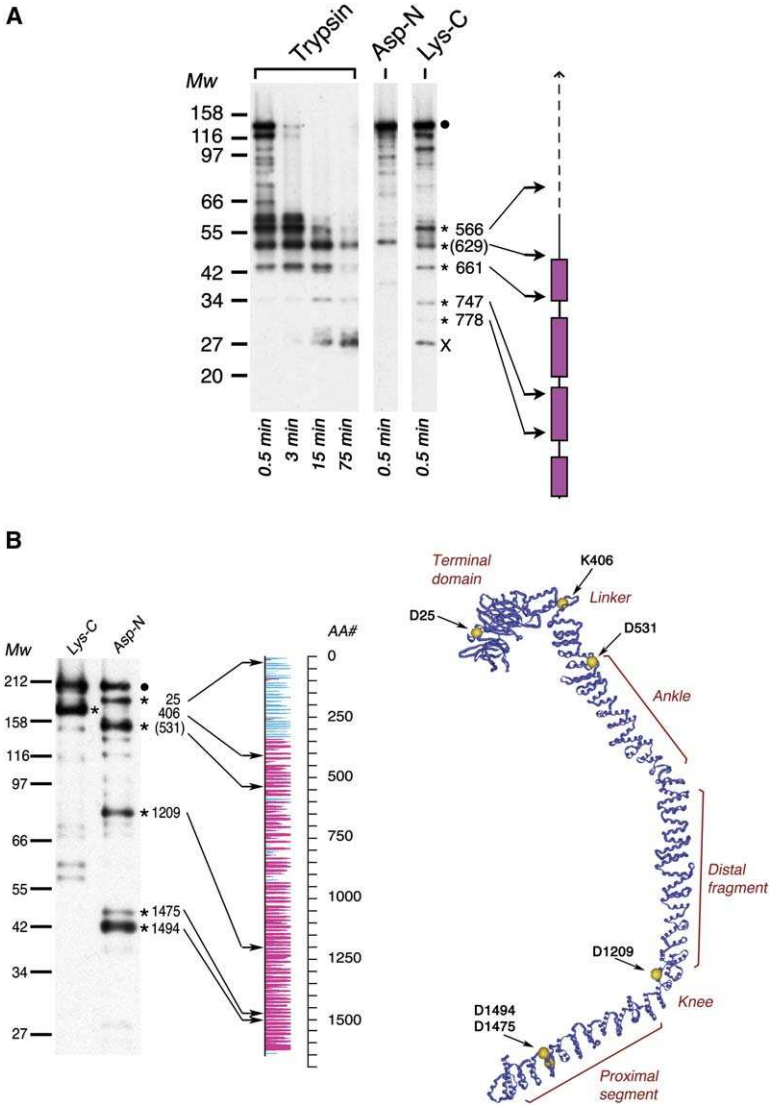

Figure 2. PAL Analysis of Nsp1 and Clathrin Heavy Chain

Immunoblots of SDS-PAGE-resolved proteolytic fragments are shown; marker protein molecular weights (Mw) are indicated in kDa at the left of each blot, full-length protein is annotated with an adjacent black dot, and the carboxy-terminal proteolytically resistant PrA fragment is indicated with an X. The position of the cleaved amino acid in each of the starred fragments is indicated to the right of the star, and was determined either by amino-terminal Edman sequencing or (when bracketed) by molecular weight estimation.

(A) PAL of Nsp1-PrA after proteolysis with trypsin, Asp-N, and Lys-C at different time points, indicated at the bottom of the blot. The carboxyl terminus of Nsp1 is diagrammed vertically to scale at the right of the gel, and shows the position of the last part of the FG repeat region (dotted line) adjacent to the coil-coiled region, consisting of four coils (purple rectangles) (Bailer et al., 2001). The positions of cleavages (found exclusively in or close to the linkers between the coil regions) are marked by arrows.

(B) Left: PAL analysis of CHC-PrA, obtained after digest with Asp-N and Lys-C for 3 min. A secondary structure prediction map is shown vertically to the right of the gel (Devos et al., 2004), where the positions of cleavages are indicated with arrows. The thin vertical line represents the primary sequence of CHC, with the scale in amino acid residues indicated. Secondary structure predictions are shown as horizontal columns to the right of the primary sequence line for β strands (β propellers; cyan) and α helices (α solenoids; magenta); the length of the columns is proportional to the confidence of the secondary structure prediction (McGuffin et al., 2000). Right: Modeled 3D structure of the yeast CHC. Major segments of the CHC are indicated in red; PAL cleavage sites, found mainly between these segments, are marked with arrows and yellow spheres.

α-helical repeats. Clathrin has also been extensively studied by limited proteolysis, providing us with useful comparative data (Kirchhausen and Harrison, 1984; Lemmon et al., 1988; Matsui and Kirchhausen, 1990). CHC, tagged with PrA, was isolated from yeast in a complex with the light chain (CLC) (data not shown). Because CLC binds CHC only at the latter's extreme carboxy-terminal proximal segment (Chen et al., 2002), the majority of the protein is not protected by this interaction and will be available for protease mapping. There are several PAL-susceptible sites on CHC (Figure 2B). We determined the cleavage sites by Edman sequencing or MW estimation, and mapped their position on a yeast clathrin structure that we modeled based on the vertebrate template (Fotin et al., 2004). A major cleavage site occurs ~50 kDa from the amino terminus, at K406, which lies in the helical zigzag linker region 330–494 between the amino-terminal β propeller and carboxy-terminal α-helical domain (ter Haar et al., 1998) (Figure 2B). This result is in agreement with earlier proteolytic data (Kirchhausen and Harrison, 1984; Lemmon et al., 1988), which detected formation of a stable 52–59 kDa (for mammalian clathrin) or 43 kDa (for yeast clathrin) amino-terminal fragment, after digest of clathrin triskelions and cages. Other proteolytically sensitive sites are all situated in the hinges between different CHC segments (D531–D557, D1209) or in loops (D25, D1494). Interestingly, D1209 and D1475 are located within an α helix, although on its exposed, solvent-accessible side and opposite the CLC binding site; perhaps CLC binding improves proteolytic accessibility to these sites.

Different Protein Tags Can Be Used for PAL

For PAL to be widely applicable, it would be advantageous if tags other than PrA could be used. The PrA tag is a good choice for PAL, because it strongly binds to the beads, the tag itself is resistant to proteases under the conditions of limited proteolysis, and PrA fragments can be easily detected by immunoblotting. In principle, any tag with similar characteristics can be used in PAL. As GFP is another ubiquitously utilized tag, we compared the proteolytic behavior of PrA- and GFP-tagged Nsp1. Like PrA, the GFP tag is proteolytically

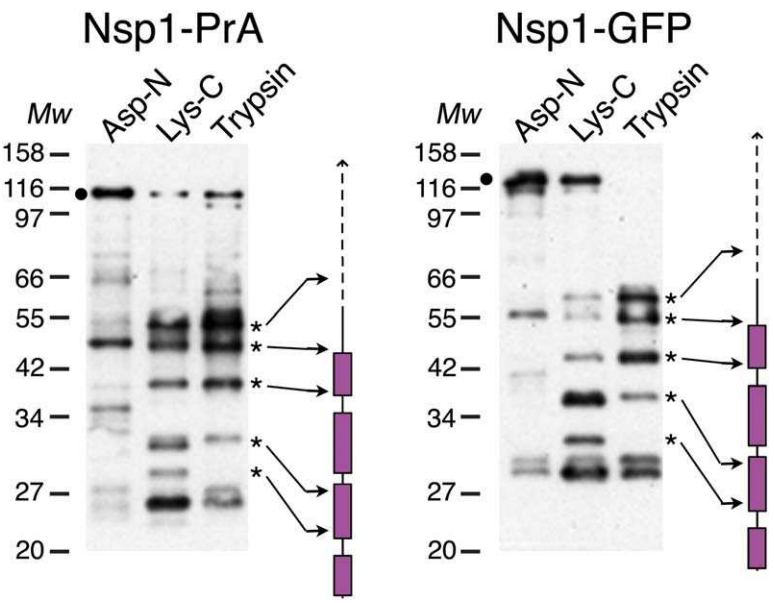

Figure 3. Comparison of Proteolytic Patterns of Nsp1-PrA and Nsp1-GFP

Both proteins were probed with the indicated proteases for 3 min. The products of proteolysis were run on the same gel and immunoblot analyses were performed with different antibodies specific to the corresponding tags (Experimental Procedures). The carboxyl terminus of Nsp1 is diagrammed vertically to scale at the right of each gel, and annotated as in Figure 2A. Though full-length Nsp1-GFP and its proteolytic fragments migrate higher than Nsp1-PrA and corresponding fragments, their fragmentation patterns are similar.

resistant using our conditions. Under extreme protease conditions (outside of the useful range) both PrA and GFP will begin to degrade (data not shown). However, should a small fraction degrade, it should fall off the antibody-bead and so not be assayed.

Because the molecular weight of the PrA tag and the GFP tag are similar (~26 kDa), it is reasonable to expect that the proteolytic fragments of the two differently tagged versions of Nsp1 should be of similar size, thus generating a similar PAL map. Figure 3 shows this is indeed the case. In addition, we have recently found that His-tagged Nsp1 purified from *Escherichia coli* and bound on TALON Dynabeads gave the same proteolytic pattern as Nsp1-PrA and Nsp1-GFP (J. Novatt and M.P.R., unpublished data). Thus, it would appear that any proteolytically resistant tag in combination with appropriate affinity beads can be used for PAL analysis.

A Combination of PAL and Comparative Protein Structure Modeling as a Tool for the Structural Characterization of Proteins

Recently, we successfully applied PAL to confirm fold predictions for the proteins of the yeast Nup84 subcomplex—the main building block of the nuclear pore complex (Devos et al., 2004). Using computational analysis, we showed that all proteins of this subcomplex contain either α solenoid or β propeller folds, or both. PAL analysis corroborated our predictions, contributing to the confidence in the fold assignments. PAL analysis revealed that proteins with β propeller fold were resistant to proteolysis, while proteins with α solenoid fold were more susceptible—though only in the regions of predicted loops. In the two-domain proteins, where an amino-terminal β propeller was connected to an α solenoid, PAL readily detected the most susceptible site within a linker between these two domains. This result confirmed that the predicted domain definitions were accurate, an important aspect of fold recognition and comparative modeling (Marti-Renom et al., 2000). Comparison of the proposed folds of nucleoporins with the structures of components of coated vesicles revealed striking similarities, which allowed us to suggest that all three major classes of coated vesicles (clathrin/adaptin, COPI and COPII complexes) can be linked together via their common architecture.

Here we investigate the structure of the Sec31 protein—a component of COPII-coated vesicles, whose X-ray structure has not been solved. Secondary structure prediction combined with fold recognition showed that Sec31 contains an amino-terminal β propeller (amino acid residues 1 to ~410) followed by a large α solenoid region, then a long stretch of 450 residues (amino acid residues 750–1100) of unusual properties (i.e., few predicted secondary structures, no satisfying fold predictions) (Figure 4). While the last ~200 carboxy-terminal residues contain α helices, it is not clear whether or not they form an α solenoid. Sec31-PrA, purified as a dimer with Sec13 protein (data not shown), was subjected to PAL analysis, which revealed several protease-accessible regions. The position of all the major cleavage sites was mapped by amino-terminal Edman sequencing (Figure 4). The first site (K423) is situated at the end of the β propeller, the second one (K524) is located at the beginning of the predicted α solenoid in a "linker" between two domains, and all others fall into the region with few predicted regular secondary structure segments. After PAL analysis was completed, fold prediction was performed again, but this time we analyzed only the fragments situated between proteolytic sites. As expected based on experience with fold assignment in general, the use of domains detected by PAL helped the fold recognition servers to predict their fold types. More specifically, (1) it increased the similarities between the top scoring folds, (2) it improved the scoring and ranking of the proposed folds, and (3) it increased the extent and quality of alignments. For example, the fold assignment for random 300 residue segments of Sec31 contained the α solenoid fold only in the fourth position among the top ten hits. In contrast, the fold assignment for the PAL-derived domain (residues 524–750) resulted in the α solenoid fold at four of the ten best scoring predictions, including the top scoring

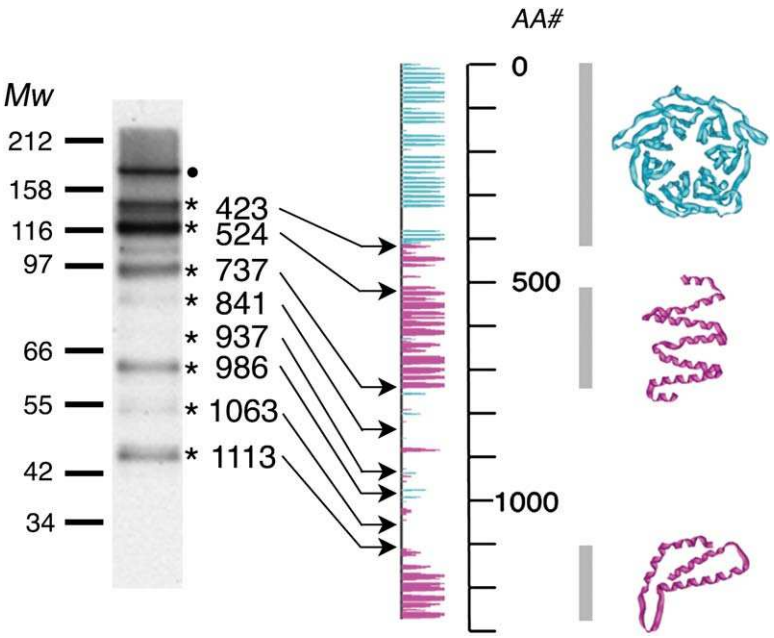

Figure 4. PAL Analysis and Fold Prediction for Sec31

Left: Immunoblot of Sec31-PrA after proteolysis with Lys-C for 3 min, annotated as in Figure 2. Right: Sites of proteolysis are indicated with arrows on a secondary structure prediction map, where β strands are shown in cyan and α helices are in magenta. Gray bars indicate the modeled regions, and their corresponding modeled β propeller (cyan) and α solenoid (magenta) structures are shown to the right of the secondary structure prediction map.

fold. The similarities between the PAL results and structural predictions strengthen the proposition that Sec31 shares the same structural features (i.e., an amino-terminal β propeller followed by an α solenoid) as the clathrin heavy chain and two members of the Nup84 subcomplex—Nup133 and Nup120 (Figure 4).

Application of PAL for the Study of Protein-Protein Interactions

To investigate the utility of PAL for the localization of protein-protein interaction interfaces, we chose the yeast Kap-α(Kap60)/Kap-β(Kap95) complex (also called the importin α/importin β complex). The Kap-α/Kap-β heterodimer targets proteins with nuclear localization signals (NLS) to the NPC and facilitates their translocation across NPCs. The X-ray crystallographic structures of both Kap-α and Kap-β have been solved, as well as the structures of the Kap-α/Kap-β heterodimer and its complex with different transport cargoes (Cingolani et al., 1999; Conti et al., 1998; Matsuura and Stewart, 2004).

We isolated Kap60-PrA from yeast cells and bound it while still on the beads to Kap95-GST, purified from *E. coli*. This complex was probed with proteases and the resulting PAL maps were compared with the PAL maps obtained after proteolysis of Kap60-PrA alone. A fragment, readily observed upon Lys-C proteolysis of Kap60-PrA alone, is absent during similar treatment of the Kap60-PrA/Kap95-GST complex (Figure 5). Edman sequencing of an analogous band, obtained after trypsin cleavage, showed that proteolysis occurred close to R54. At the same time, two lower minor bands became more accessible for the proteolysis in the protein heterodimer. Thus, the major changes in Kap60 proteolytic sensitivity were observed at the extreme amino-terminal portion of the protein. This observation is in agreement with numerous biochemical and structural data, as well as limited proteolysis probing, which defines the ~50 amino-terminal residues of Kap60 as the Kap-β binding domain, a region protected in the Kap60/Kap95 complex (also called the importin β binding domain [IBB]) (Cingolani et al., 1999, 2000; Conti et al., 1998; Gorlich et al., 1996; Matsuura and Stewart, 2004). Hence, PAL is potentially a rapid technique to map protein-protein interaction sites.

Discussion

We report here a new variation of a limited proteolysis method, which we call protease accessibility laddering (PAL). PAL has several important advantages when compared with "classical" limited proteolysis approaches.

First, PAL does not necessarily require expression and purification of proteins from heterologous expression systems. In this study, we demonstrate that genomically tagged yeast proteins (i.e., expressed at their natural levels) can be readily isolated from their natural (yeast) environment, which generally ensures that they are studied in their properly folded state. This is especially important for large, multidomain proteins, whose purification in soluble form from bacteria can be challenging due to protein misfolding. However, heterologously expressed proteins with appropriate tags can also be successfully tested with PAL; their PAL maps can be compared with those of natively expressed, tagged versions of the same protein (i.e., isolated from their normal parent organism) to confirm proper folding of the heterologously expressed version.

Second, use of magnetic beads in PAL provides a gentle protein isolation procedure and assures a high yield of purification. In addition, because PAL requires only small amounts of protein (<1 μg per reaction), many proteolytic conditions can be tested at the same time. Another advantage of the beads is that binding of tagged proteins to the beads is fast (due to the large and exposed bead surface area). This highly exposed surface area, combined with the high speed of magnetic separation, also ensures the immediate removal of enzyme and unbound proteins after reaction; hence, the entire PAL

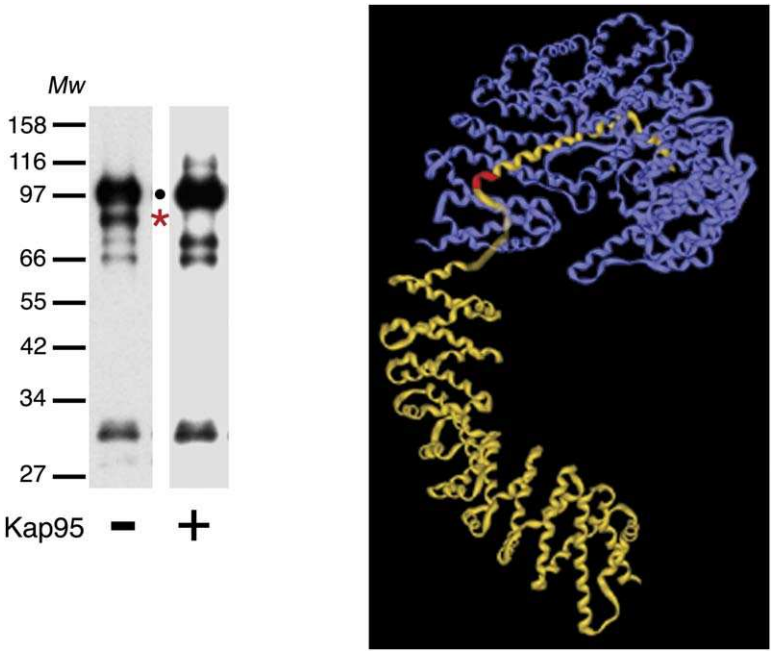

Figure 5. PAL Study of the Kap60-Kap95 Interaction

Left: Immunoblot of Kap60-PrA after proteolysis with Lys-C for 30 s in the absence or presence of Kap95-GST. The protected fragment is indicated with a red star. Right: Modeled structure of Kap95 (blue) (PDB code 1QGK) (Cingolani et al., 1999) docked to Kap60 (yellow) (PDB code 1BK5) (Conti et al., 1998). Kap60 amino acid residue R54, adjacent to the region of Kap60 protected by Kap95, is shown in red.

procedure is rapid. Magnetic beads obviously work not only for yeast but also for any other cell extract. For example, isolation of functional GFP-tagged proteins from HeLa cell extracts on magnetic beads is straightforward (Cristea et al., 2005). We believe that the PAL method should be equally applicable to other such protein systems.

Third, proteolytic fragments in PAL are detected by immunoblot analysis with the appropriate antibodies against the tag, which are generally inexpensive and commercially available. There is a special advantage to using carboxy-terminal tags, because they enable amino-terminal Edman sequencing; this ability would be lost if tags were fused at the protein N terminus instead. Mass spectrometric analysis can also be applied for the identification of the exact cleavage sites (Cohen and Chait, 2001). However, the SDS-PAGE immunoblot readout technique is rapid and robust, and can be performed on nanogram amounts of proteolytic products irrespective of the molecular mass of the protein under study.

In this study, we show that such widely used tags as PrA, GFP, and 6×His can be successfully applied in PAL. Any proteolytically resistant tag, in combination with the corresponding tag binding magnetic beads, should be appropriate for PAL studies. The great advantage of the GFP tag is that it can also serve simultaneously for localization, structural, and functional experiments. In addition, a large set of yeast GFP-tagged strains, produced in the E. Shea laboratory (Huh et al., 2003), is now commercially available and could immediately be probed by PAL.

Like other proteolytic mapping techniques, PAL defines the most stable fragments in a protein by trimming the flexible regions between domains (Fontana et al., 2004; Hubbard, 1998). These fragments are potentially the best candidates for crystallization studies. Combining PAL with bioinformatics methods, such as secondary structure predictions, fold recognition, and comparative protein structure modeling, reveals another power of the method. We have shown that (1) PAL can readily scan the accessible sites of proteins of known structure, (2) PAL confirms fold recognitions and sequence-structure alignments, adding confidence to the predictions, and (3) similarities in PAL results confirm a structural relationship between potentially related proteins.

PAL is a rapid and robust technique to proteolytically map natively expressed proteins of any size, purified from their normal environment. PAL is readily applicable to a high-throughput experimental format, and we believe that PAL will be of great utility to both small- and large-scale proteomics and structural genomics studies.

Experimental Procedures

Yeast Strains and Materials

Genomic tagging of CHC and Sec31 with PrA in W303 *Saccharomyces cerevisiae* strain was performed as in Rout et al. (2000). Nsp1-PrA and Kap60-PrA were from Rout et al. (2000), and Nsp1-GFP was from Dilworth et al. (2001). Kap95-GST, purified from *E. coli*, was kindly provided by Jacklyn Novatt (Rout laboratory). The following materials were used in this study: 2.8 μm Dynabeads M-270 Epoxy (143.02; Dynal, Oslo, Norway); rabbit IgG (55944; ICN Biochemicals, Costa Mesa, CA); protease inhibitor cocktail (P-8340; Sigma, St. Louis, MO); Asp-N, Lys-C, and trypsin (11420488001, 11420429001, 11418025001; Roche Diagnostics, Indianapolis, IN); HRP-rabbit IgG (011-0303-003; Jackson ImmunoResearch, West Grove, PA); anti-GFP antibody (11814460001; Roche Diagnostics); and anti-mouse IgG-HRP (NA931V; Amersham Biosciences UK Limited, Little Chalfont, Buckinghamshire, UK). We also used a ball mill (MM301; Retsch, Haan, Germany), Polytron (Kinematica, Littau-Luzerne, Switzerland), 4%–12% bis-Tris or 4%–20% Tris-glycine gels and PVDF membrane 0.2 μm pore size (Invitrogen, Carlsbad, CA), as well as nitrocellulose membrane 0.45 μm pore size (Bio-Rad Laboratories, Hercules, CA).

Protease Accessibility Laddering

Yeast cells carrying PrA- or GFP-tagged proteins were grown and harvested as described previously (Rout et al., 2000). Cell pellets were frozen in liquid nitrogen and homogenized to a fine powder

in a ball mill, and cooled with liquid nitrogen. Half a gram of cell powder was normally used for one proteolytic reaction. The cell powder was thawed on ice, ten volumes of extraction buffer were added to cells, and homogenized at 4ºC with a Polytron. To isolate an individual protein, we used extraction buffer EB (20 mM K/HEPES [pH 7.4], 1.0% Triton X-100, 0.5% sodium deoxycholate, 0.3% sodium N-lauroyl-sarcosine, 0.1 mM $MgCl_2$, 1 mM DTT, 1:500 protease inhibitor cocktail). Usually, we obtain 1–3 μg of pure protein from 0.5 g of cells. For protein complexes, we used extraction buffer TBT (20 mM K/HEPES [pH 7.4], 110 mM KAc, 2 mM $MgCl_2$, 0.1% Tween 20, 1 mM DTT, 1:500 protease inhibitor cocktail), containing 1% Triton X-100 and 75 mM NaCl. The cell lysate was clarified by centrifugation (2000 × g for 15 min, 4ºC). Magnetic beads were conjugated to rabbit IgG or anti-GFP antibody (kindly provided by R.W., I. Cristea, and M.P.R.) (Cristea et al., 2005) according to the manufacturer's instructions. The beads were added to the extract to a ratio of about 8×10^8 beads per gram of cells. After incubation for 1 hr at 4ºC with slow rotation, the beads were magnetically recovered (we recommend using neodymium iron boron magnets), washed five times with 1 ml of 20 mM K/HEPES (pH 7.4), 1 mM EDTA, 0.1% Triton X-100, 0.05% sodium deoxycholate, 0.03% sodium N-lauroyl-sarcosine, or with TBT containing 1 mg/ml of heparin, or with TBT containing 200 mM $MgCl_2$ (for Kap60-PrA purification), and then resuspended in 50 μl of reaction buffer. Reaction buffers were prepared according to the manufacturer's specifications and were the following: for Asp-N: 50 mM sodium phosphate (pH 8.0), 0.01% SDS; for Lys-C: 25 mM Tris-HCl (pH 8.5), 1 mM EDTA, 0.01% SDS; for trypsin: 100 mM Tris-HCl (pH 8.5), 0.01% SDS. Protease was added to give a weight ratio of 1:200 of protease to the tagged protein. After incubation with mild agitation in the shaker at different time points (usually 30 s, 3 min, 15 min, 75 min) at 37ºC, beads were magnetically harvested, washed once with 1 ml of 0.1 M NH_4OAc, 0.1 mM $MgCl_2$, 0.02% Tween 20, and tagged fragments were eluted with 1 ml of 0.5 M NH_4OH containing 0.5 mM EDTA. The eluant was vacuum dried, resuspended in SDS-PAGE sample buffer, and separated on 4%–12% bis-Tris or 4%–20% Tris-glycine gels. Proteins were then transferred electrophoretically to a nitrocellulose membrane and probed with HRP-IgG for PrA detection or with anti-GFP antibody, followed by anti-mouse HRP-IgG, for GFP detection. For analysis by amino-terminal Edman sequencing, proteins were transferred to a PVDF membrane (Fernandez et al., 1994).

Bioinformatics Methods

Yeast clathrin homologs were detected by Psi-blast (Altschul et al., 1997) for five iterations with default parameters. Multiple sequence alignment was built with T-coffee (Poirot et al., 2004) with default parameters. The yeast clathrin was modeled based on the bovine chain (Fotin et al., 2004) using comparative protein structure modeling by satisfaction of spatial restraints as implemented in MODELLER-8 (http://salilab.org/modeller; Sali and Blundell, 1993) (template, Protein Data Bank [PDB] code 1IX4, Z score −19.77). Secondary structure segments of clathrin and Sec31 were predicted from sequence only by the PSIPRED server (McGuffin et al., 2000). Fold recognition used the same server and HHSearch (Soding, 2005) (β propeller: template, PDB code 1GXR, Z score −7.8; α solenoid: template, PDB code 1QQE, Z score −4.4; carboxyl terminus: template, PDB code 1UPK, Z score −4.5).

Acknowledgments

We gratefully acknowledge the valuable assistance of Joe Fernandez, Jacklyn Novatt, and Ileana Cristea throughout the course of this study as well as members of the Rout and Chait laboratories for sharing their protocols and reagents. Protein sequence analysis was provided by The Rockefeller University Protein/DNA Technology Center, which is supported in part by NIH shared instrumentation grants and by funds provided by the U.S. Army and Navy for purchase of equipment. This work was supported by an Irma T. Hirschl Career Scientist Award, a Sinsheimer Scholar Award, and a grant from the Rita Allen Foundation to M.P.R., and grants from the National Institutes of Health to M.P.R. (GM062427), B.T.C. (RR00862), B.T.C. and M.P.R. (CA89810), M.P.R., B.T.C., A.S. (RR022220), and A.S. (GM54762, GM62529). A.S. was also supported by the Sandler Family Foundation, an IBM SUR grant, and an Intel computer hardware gift.

Received: December 20, 2005
Revised: February 1, 2006
Accepted: February 2, 2006
Published: April 11, 2006

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Bailer, S.M., Balduf, C., and Hurt, E. (2001). The Nsp1p carboxy-terminal domain is organized into functionally distinct coiled-coil regions required for assembly of nucleoporin subcomplexes and nucleocytoplasmic transport. Mol. Cell. Biol. *21*, 7944–7955.

Chen, C.Y., Reese, M.L., Hwang, P.K., Ota, N., Agard, D., and Brodsky, F.M. (2002). Clathrin light and heavy chain interface: α-helix binding superhelix loops via critical tryptophans. EMBO J. *21*, 6072–6082.

Cingolani, G., Petosa, C., Weis, K., and Muller, C.W. (1999). Structure of importin-β bound to the IBB domain of importin-α. Nature *399*, 221–229.

Cingolani, G., Lashuel, H.A., Gerace, L., and Muller, C.W. (2000). Nuclear import factors importin α and importin β undergo mutually induced conformational changes upon association. FEBS Lett. *484*, 291–298.

Cohen, S.L., and Chait, B.T. (2001). Mass spectrometry as a tool for protein crystallography. Annu. Rev. Biophys. Biomol. Struct. *30*, 67–85.

Cohen, S.L., Ferre, D.A.A.R., Burley, S.K., and Chait, B.T. (1995). Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. Protein Sci. *4*, 1088–1099.

Conti, E., Uy, M., Leighton, L., Blobel, G., and Kuriyan, J. (1998). Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α. Cell *94*, 193–204.

Cristea, I.M., Williams, R., Chait, B.T., and Rout, M.P. (2005). Fluorescent proteins as proteomic probes. Mol. Cell. Proteomics *4*, 1933–1941.

Denning, D.P., Patel, S.S., Uversky, V., Fink, A.L., and Rexach, M. (2003). Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. Proc. Natl. Acad. Sci. USA *100*, 2450–2455.

Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B.T., Sali, A., and Rout, M.P. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. PLoS Biol. *2*, e380.

Dilworth, D.J., Suprapto, A., Padovan, J.C., Chait, B.T., Wozniak, R.W., Rout, M.P., and Aitchison, J.D. (2001). Nup2p dynamically associates with the distal regions of the yeast nuclear pore complex. J. Cell Biol. *153*, 1465–1478.

Fernandez, J., Andrews, L., and Mische, S.M. (1994). An improved procedure for enzymatic digestion of polyvinylidene difluoride-bound proteins for internal sequence analysis. Anal. Biochem. *218*, 112–117.

Fontana, A., de Laureto, P.P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004). Probing protein structure by limited proteolysis. Acta Biochim. Pol. *51*, 299–321.

Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S.C., Kirchhausen, T., and Walz, T. (2004). Molecular model for a complete clathrin lattice from electron cryomicroscopy. Nature *432*, 573–579.

Gorlich, D., Henklein, P., Laskey, R.A., and Hartmann, E. (1996). A 41 amino acid motif in importin-α confers binding to importin-β and hence transit into the nucleus. EMBO J. *15*, 1810–1817.

Hubbard, S.J. (1998). The structural aspects of limited proteolysis of native proteins. Biochim. Biophys. Acta *1382*, 191–206.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Kirchhausen, T., and Harrison, S.C. (1984). Structural domains of clathrin heavy chains. J. Cell Biol. *99*, 1725–1734.

Lemmon, S., Lemmon, V.P., and Jones, E.W. (1988). Characterization of yeast clathrin and anticlathrin heavy-chain monoclonal antibodies. J. Cell. Biochem. *36*, 329–340.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. *29*, 291–325.

Matsui, W., and Kirchhausen, T. (1990). Stabilization of clathrin coats by the core of the clathrin-associated protein complex AP-2. Biochemistry *29*, 10791–10798.

Matsuura, Y., and Stewart, M. (2004). Structural basis for the assembly of a nuclear export complex. Nature *432*, 872–877.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404–405.

Neurath, H. (1980). Limited proteolysis, protein folding and physiological regulation. In Protein Folding, R. Jaenicke, ed. (Amsterdam-New York: Elsevier/North Holland Biomedical Press), pp. 501–504.

Pedigo, S., and Shea, M.A. (1995). Quantitative endoproteinase GluC footprinting of cooperative Ca2+ binding to calmodulin: proteolytic susceptibility of E31 and E87 indicates interdomain interactions. Biochemistry *34*, 1179–1196.

Poirot, O., Suhre, K., Abergel, C., O'Toole, E., and Notredame, C. (2004). 3DCoffee@igs: a web server for combining sequences and structures into a multiple sequence alignment. Nucleic Acids Res. *32*, W37–W40.

Polverino de Laureto, P., Taddei, N., Frare, E., Capanni, C., Costantini, S., Zurdo, J., Chiti, F., Dobson, C.M., and Fontana, A. (2003). Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis. J. Mol. Biol. *334*, 129–141.

Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B.T. (2000). The yeast nuclear pore complex: composition, architecture, and transport mechanism. J. Cell Biol. *148*, 635–651.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779–815.

Scaloni, A., Miraglia, N., Orru, S., Amodeo, P., Motta, A., Marino, G., and Pucci, P. (1998). Topology of the calmodulin-melittin complex. J. Mol. Biol. *277*, 945–958.

Soding, J. (2005). Protein homology detection by HMM-HMM comparison. Bioinformatics *21*, 951–960.

ter Haar, E., Musacchio, A., Harrison, S.C., and Kirchhausen, T. (1998). Atomic structure of clathrin: a β propeller terminal domain joins an α zigzag linker. Cell *95*, 563–573.

Ybe, J.A., Brodsky, F.M., Hofmann, K., Lin, K., Liu, S.H., Chen, L., Earnest, T.N., Fletterick, R.J., and Hwang, P.K. (1999). Clathrin self-assembly is mediated by a tandemly repeated superhelix. Nature *399*, 371–375.

Zhong, M., Lin, L., and Kallenbach, N.R. (1995). A method for probing the topography and interactions of proteins: footprinting of myoglobin. Proc. Natl. Acad. Sci. USA *92*, 2111–2115.