Recombinant expression and domain structure of the Rna1 protein from *Schizosaccharomyces pombe*

Jörg Haberland, Volker Gerke*

Max Planck Institute for Biophysical Chemistry, Department of Biochemistry, PO Box 2841, D-37018 Göttingen, Germany

Received 24 November 1994

Abstract The amino acid sequence of Rna1p, a yeast protein implicated in the maturation and/or nucleocytoplasmic transport of RNA, is characterised by the presence of eight leucine-rich repeats (LLRs) as well as two intervening repeats of a different type and a highly acidic C-terminal region. Limited proteolysis of purified Rna1p expressed recombinantly in bacteria reveals that the C-terminal extension but not the region containing the two types of repeats is highly accessible to proteolytic attack and that the C-terminal region most likely harbours (a) low affinity Ca^{2+}-binding site(s). These results are indicative of the domain structure of the Rna1p molecule, with the repeats and the C-terminal region being accessible for different interactions.

Key words: Leucine-rich repeat; Nucleocytoplasmic transport; *RNA1* mutant; RNA processing

1. Introduction

The *rna* mutants of the yeast *Saccharomyces cerevisiae* are defined as a set of thermosensitive mutant strains in which the production of RNA is affected after the shift to the restrictive temperature [1]. While the *rna2–11* mutants are impaired in the proper processing of pre-mRNA (for review see [2]) the *rna1* mutant is an exception as it shows pleiotropic defects in the production of all major classes of RNA and in the nucleocytoplasmic transport of mRNA (for review see [3]).

Although the phenotypic consequences of a conditionally lethal mutation in the *RNA1* gene have been described in detail, the precise function of its product is not known. Likewise, the structural and biochemical information on the Rna1 protein is so far limited to the respective cDNA and/or genomic sequences from *S. cerevisiae* [4], *Schizosaccharomyces pombe* [5], mice [6] and *Caenorhabditis elegans* [7], and to a purification protocol for *S. pombe* Rna1p [5]. Internal homology searches revealed that the deduced amino acid sequences of all Rna1 proteins show a pronounced repeat structure with a leucin-rich motif of 29 amino acids repeated eight times along the polypeptide chain [5,6,8]. These leucine-rich repeats (LRRs), which are characterised by regularly spaced hydrophobic, predominantly leucine residues, are present in a number of proteins otherwise diverse in function and cellular localisation, e.g. the RNase inhibitor [9], adenylyl cyclase [10], the *toll* membrane receptor of *Drosophila melanogaster* [11], and the product of the disease resistance gene *RPS2* of *Arabidopsis thaliana* [12]. It is generally assumed that LRR motifs are involved in mediating highly specific protein–protein interactions, although candidates for LRR protein ligands have only been identified in a few cases, e.g. the *dis2* protein phosphatase which physically interacts with the LRR protein *sds22* of *S. pombe* [13] and RNase which forms a tight stoichiometric complex with the RNase inhibitor [14].

The recently determined crystal structure of the porcine RNase inhibitor provided the first detailed information on the three-dimensional folding of LRRs [15]. Here, the individual repeats constitute β–α structural units with the side chains of the conserved leucines or other aliphatic residues forming the hydrophobic core of the protein. The entire protein has the non-globular shape of a horseshoe with the α-helices of the LRRs aligning on the outer circumference and the β-strands forming a curved parallel β-sheet on the inner circumference. Although the lengths of the repeats and the spacing of some of the hydrophobic residues vary slightly between the repeats in different LRR proteins it seems likely that similar β–α structures occur in all proteins containing LRRs [16]. It is, however, impossible to deduce the general folding of LRR proteins from the structure of the RNase inhibitior as the number of individual repeats differs drastically between different proteins. Moreover, while the RNase inhibitor is composed exclusively of LRRs, other members of this group of proteins contain additional sequences (for review see [16]). In the case of the Rna1 protein, the LRRs are interrupted after the first and third repeat by segments of 36 amino acid residues also showing homologies with one another. In addition, a unique C-terminal extension differing in length between the Rna1 proteins from yeast, *C. elegans* and mouse follows the last LRR [5–7]. To analyse whether these non-LRR sequences/domains in Rna1p are particularly accessible in solution, thus enabling them to engage in specific interactions and/or functions, we subjected recombinantly expressed Rna1p from *S. pombe* to a series of limited proteolysis studies. Immunoblotting with domain-specific antibodies and N-terminal protein sequence analysis of proteolytic fragments revealed that the C-terminal extension of Rna1p is highly accessible to different proteases and most likely contains (a) low affinity Ca^{2+}-binding site(s).

2. Materials and methods

2.1. Expression of Rna1p in E. coli and purification of the protein

A *Bsp*M1–*Xho*1 fragment of the *S. pombe rna1* gene [5] served as the DNA template for the amplification of the complete Rna1p cDNA by PCR. Oligonucleotide O1 (5′-CTTTCAACAACGAATTCCATGTC-GCGT-3′), which introduced an *Eco*RI cleavage site upstream of the ATG start codon, was used as the sense primer, whereas oligonucleotide O2 (5′-CGGAAACAGCGCTGCAGGGTTCGAGAGGG-3′), which contained a *Pst*I cleavage site downstream of the stop codon, represented the antisense primer. The PCR product was gel-purified, cut with *Eco*RI and *Pst*I, and cloned into the appropriately linearised

*Corresponding author. *Present address:* Clinical Research Group for Endothelial Cell Biology, University of Münster, Von-Esmarch-Str. 56, D-48149 Münster, Germany. Fax: (49) (251) 836 748.

SSDI 0014-5793(94)01353-5

bacterial expression vector pkk223-3 (Pharmacia). The resulting plasmid, pkkRna1, was used to transform *E. coli* JM 105 cells, and clones containing the plasmid were selected by growth on LB containing 0.15 mg/ml ampicillin [17]. To verify that no mutations were introduced by the PCR reaction, the *rna1* cDNA insert in pkkRna1 was sequenced completely following the dideoxy method [18]. The expression of recombinant Rna1p in *E. coli* JM 105 cells containing the pkkRna1 plasmid was verified by analysing total protein extracts of bacteria, which were grown overnight in LB containing 0.15 mg/ml ampicillin and 1 mM IPTG, in immunoblots using a Rna1p-specific antibody [5].

E. coli cells containing pkkRna1 were grown at 37°C in 500 ml LB containing 0.15 mg/ml ampicillin and expression of the *rna1* gene was induced by the addition of 1 mM IPTG at late-log phase. Cells were collected by centrifugation (20 min, 5000 × *g*, 4°C) and washed once in ice-cold phosphate-buffered saline. The cell pellet was then resuspended in 7 ml buffer A (30 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM M_gCl_2, 1 mM DTT, 1 mM PMSF) and homogenised using a French press. Three vols. of buffer A were added to the homogenate and the resulting suspension was clarified by centrifugation (20 min at 10,000 × *g*, followed by 60 min at 100,000 × *g*, 4°C). Ammonium sulfate was added to the lysate to 60% saturation and precipitated proteins were removed by centrifugation. The resulting supernatant was dialysed against buffer B (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM DTT, 1 mM PMSF) and chromatographed on a DE-52 (Whatman) column which was developed with a linear gradient of 0 to 1 M NaCl in buffer B. Fractions containing Rna1p were identified by immunoblotting with a Rna1p-specific antibody, pooled, dialysed against buffer B, and applied to a Mono Q column connected to a fast-performance liquid chromatography (FPLC) system (Pharmacia). The column was washed with buffer B containing 250 mM NaCl and then developed with a linear salt gradient (0.25–1 M NaCl) in buffer B. Rna1p-containing fractions eluting at approximately 0.5 M NaCl were pooled, dialysed against buffer B, and stored at 4°C. For long-term storage the protein was dialysed against buffer B containing 50% glycerol and then kept at −20°C.

2.2. Limited proteolysis

α-Chymotrypsin and endoproteinase Asp-N treatment of purified Rna1p was carried out in buffer B with the protein concentration adjusted to 0.5 mg/ml and at an enzyme-to-substrate ratio of 1:100 and 1:150, respectively. Samples were incubated at room temperature and the reaction was stopped after the time indicated by boiling in SDS sample buffer [19]. Reaction products were separated in SDS-12.5%-polyacrylamide gels and either stained with Coomassie blue or subjected to immunoblotting [20]. To assess a potential effect of Ca^{2+}, equivalent digestions were carried out in the presence of 1, 2, 5 and 10 mM $CaCl_2$, respectively.

2.3. Antibodies

Synthetic peptides corresponding in sequence to amino acids 36–49 and 60–73 of *S. pombe* Rna1p, respectively, but containing an additional cysteine at the N-terminus, were coupled to ovalbumin using the bifunctional cross-linker *m*-maleimidobenzoyl-*N*-hydrosuccinimide ester (MBS; Pierce) [21]. Rabbits were injected with the peptide–ovalbumin conjugates as described [5] and peptide-specific antibodies were purified from the serum by affinity chromatography on the immobilised peptide [21]. The antibody directed against the C-terminal Rna1p peptide (amino acids 374–386 of *S. pombe* Rna1p) has been described [5]. For immunoblotting the affinity-purified antibodies were employed at a concentration of 2 μg/ml and visualised using peroxidase-coupled secondary antibodies (Dako) and a chemoluminescence-based detection system (Amersham-Buchler).

3. Results

3.1. Expression and purification of recombinant Rna1p

To obtain quantities of *S. pombe* Rna1p sufficient for a detailed biochemical analysis the cDNA encoding the protein was amplified by PCR and cloned into the bacterial expression vector pkk223 under the control of the *tac* promoter. Expression of the protein was induced by cultivating the transformed bacteria in the presence of IPTG as revealed by immunoblot analysis of a total bacterial lysate with a Rna1p-specific antibody (not shown).

Purification of the recombinant Rna1p was achieved by employing in a modified form a protocol developed for the isolation of the endogenous Rna1p from *S. pombe* cells [5]. Because of the highly acidic nature of the protein (the isoelectric point for *S. pombe* Rna1p is 4.3) ammonium sulfate fractionation

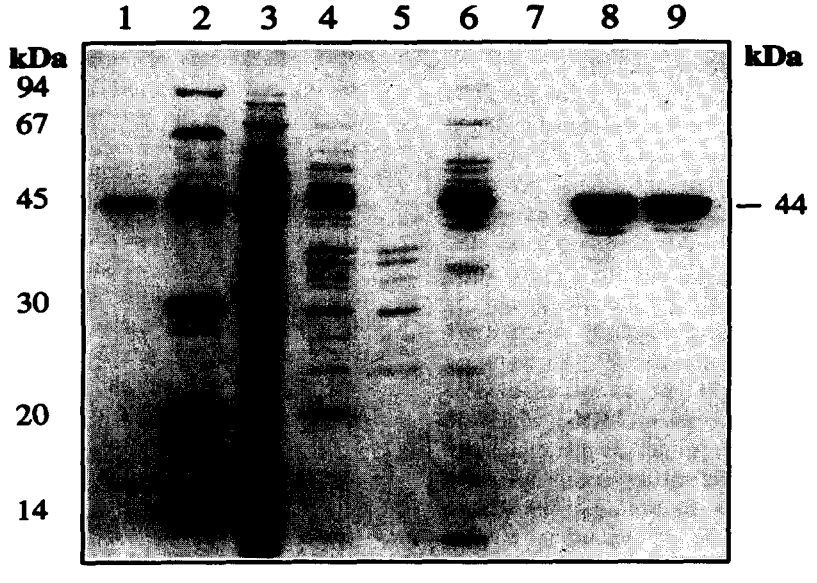

Fig. 1. Purification of *S. pombe* Rna1p expressed in *E. coli*. The purification was carried out as described in section 2 and proteins present in the different fractions were analysed by SDS-PAGE. *E. coli* cells containing the expression plasmid pkkRna1 were lysed using a French press and soluble proteins (lane 3) were fractionated by ammonium sulfate precipitation. Proteins soluble in the presence of 60% ammonium sulfate (lane 4) were dialysed against buffer B and chromatographed on a DE-52 column. Unbound material is shown in lane 5 whereas the Rna1p fraction, eluting at approximately 0.5 M NaCl, is given in lane 6. Final chromatography was performed on a Mono Q column and yielded highly purified Rna1p (lanes 8 and 9 showing a peak fraction and the entire Rna1p pool after Mono Q chromatography, respectively; the flow-through fraction of the Mono Q column is given in lane 7). Authentic Rna1p purified from *S. pombe* and molecular weight markers are shown in lanes 1 and 2, respectively.

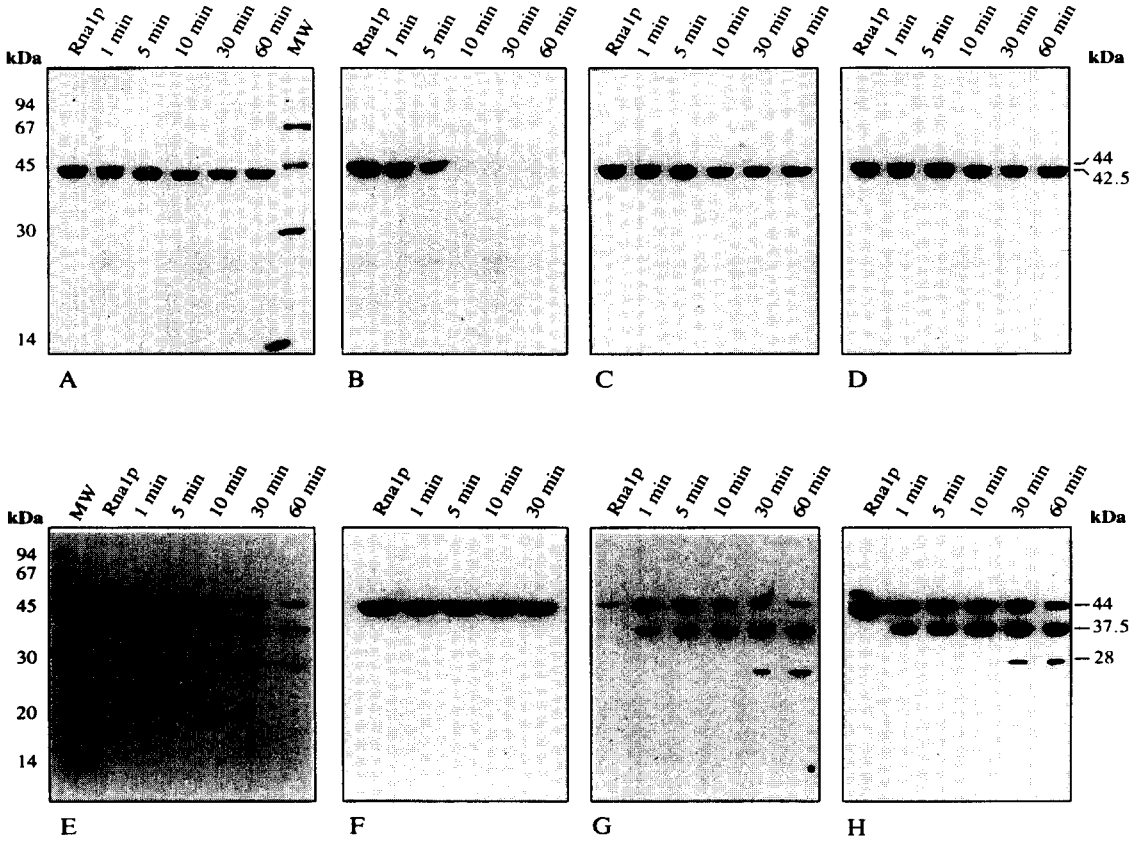

Fig. 2. Limited proteolysis of purified Rna1p. The protein was treated with endoproteinase Asp-N (upper panels, A–D) or α-chymotrypsin (lower panels, E–H) for the times indicated. Reactions were stopped by boiling the samples in SDS-PAGE sample buffer. The products of the limited proteolysis were seperated in SDS-12.5%-polyacrylamide gels and either stained with Coomassie blue (A,E) or blotted to nitrocellulose and probed with antibodies directed against peptide P1 (C,G), peptide P2 (D,H) or the C-terminal Rna1p-peptide P3 (B,F). Molecular weight markers with the apparent masses indicated are given in lane MW. Rna1p designates the unproteolysed protein. Note that both enzymes generate relatively stable fragments of 42.5 and 37.5 kDa, respectively, which have lost the C-terminal peptide.

followed by two ion exchange chromatography steps led to the preparation of Rna1p which was at least 95% pure as judged by SDS-PAGE (Fig. 1). Following this procedure approximately 0.2 mg Rna1p could be purified from a 500 ml bacterial culture. The pure protein was recognised by an antibody directed against a synthetic peptide corresponding in sequence to the C-terminal 12 amino acid residues of *S. pombe* Rna1p (not shown). Moreover, when analysed in analytic gel filtration and sucrose velocity gradients the recombinant Rna1p has the same hydrodynamic properties as authentic Rna1p from

S. pombe (not shown). These data indicate that Rna1p expressed in bacteria most likely has the same general folding as the authentic protein synthesised in *S. pombe* cells.

3.2 Probing of the domain structure of Rna1p by limited proteolysis

The primary structure of both *S. pombe* and *S. cerevisiae* Rna1p is characterised by the presence of three principle sequence elements. Eight leucine-rich repeats, each 29 amino residues in length, two intervening repeats of 36 amino acids which interrupt the otherwise continuous LRRs after the first and the third LRR, and a region C-terminally flanking the LLRs which is characterised by a stretch of acidic amino acids followed by a short amphiphatic α-helix at the C-terminus [5]. To analyse whether these sequence elements also form distinct structural domains in the folded molecule we subjected purified Rna1p to a series of limited proteolysis studies and characterised the resulting fragments by immunoblotting with three different peptide-specific antibodies. The sequence segments recognised by these antibodies reside in the first LRR (amino acids 36–49, peptide P1), the first intervening repeat (amino acids 60–73, peptide P2), and the C-terminal helix (amino acids 374–386, peptide P3), respectively.

We first employed endoproteinase Asp-N which has a specificity for aspartate and thus was the preferable enzyme to characterise the accessibility of the aspartate-rich C-terminal extension. The time-course shown in Fig. 2A reveals that Asp-N cleavage generates a stable 42.5 kDa fragment, which is already apparent after 1 min of digestion. Five minutes of proteolysis converts the intact Rna1p almost completely into the 42.5 kDa product which then remains stable for at least 1 h. Immunoblotting with the three antibodies reveals that only anti-P1 and anti-P2 but not anti-P3 recognise the 42.5 kDa product of the limited Asp-N digestion (Fig. 2B–D). This analysis clearly indicates that Asp-N cleavage removes the very C-terminal end of the protein, with the hydrolysis most likely occurring in the stretch of acidic amino acids present in the C-terminal extension. A highly similar proteolysis pattern is obtained when endoproteinase Glu-C, an enzyme cleaving C-terminally to both aspartate and glutamate, is employed (not shown).

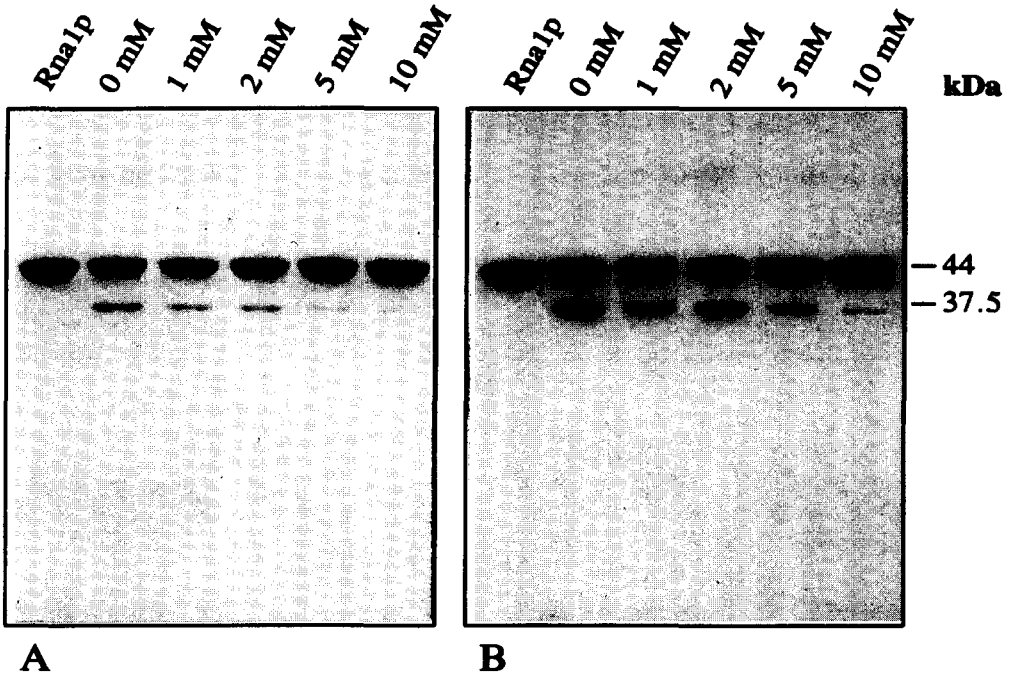

Fig. 3. Effect of Ca^{2+} on the stability of Rna1p towards α-chymotrypsin. Purified Rna1p was subjected to 30 min α-chymotrypsin treatment as outlined in section 2. The reaction was carried in the presence of 0, 1, 2, 5 or 10 mM Ca^{2+} and products were analysed by SDS-PAGE and Coomassie staining (A) or immunoblotting with the antibody directed against peptide P1. Note that the stability of the protein against chymotryptic attack is increased in the presence of elevated Ca^{2+} concentrations.

Limited proteolysis with α-chymotrypsin, which cleaves C-terminally to aromatic and also other hydrophobic amino acids, yields a major fragment of around 37.5 kDa (Fig. 2E). This product is already apparent after 1 min of digestion and remains relatively stable throughout the next 60 min of α-chymotrypsin treatment. Only upon prolonged incubation with the enzyme does a 28 kDa fragment becomes visible in the SDS-PAGE analysis shown in Fig. 2E. Both chymotryptic fragments react with the anti-P1 and anti-P2 but not the anti-P3 antibodies indicating that they represent C-terminally truncated versions of the protein (Fig. 2F–H). To corroborate this conclusion the fragments generated after 30 min of chymotrypsin treatment were separated by SDS-PAGE, blotted to PVDF membrane and subjected to N-terminal sequence analysis. All polypeptides present, i.e. the intact 44 kDa protein as well as the 37.5 kDa and 28 kDa chymotryptic fragments, start with the sequence SRSF corresponding to amino acid residues 2–5 of the Rna1p sequence encoded by the *S. pombe* cDNA. This indicates that the starting methionine residue is removed post-translationally in *E. coli* and that the resulting N-terminal serine remains otherwise unmodified. In contrast, authentic Rna1p isolated from *S. pombe* is N-terminally blocked to automated Edman degradation (data not shown). Taken together these data show that α-chymotrypsin generates stable fragments lacking a part of the C-terminal but not the N-terminal sequence. Moreover, the size of the major chymotryptic fragment, 37.5 kDa, indicates that the cleavage has occurred in the C-terminal extension close to the last LRR resulting in the loss of an approximately 7 kDa portion comprising most or all of the highly acidic stretch and the C-terminal helix.

Previous studies have indicated the presence of (a) low affinity Ca^{2+}-binding site(s) in *S. pombe* Rna1p [5]. To study a potential effect of the binding of Ca^{2+} on the stability of Rna1p, we performed the proteolysis described above in the presence of varying Ca^{2+} concentrations. Fig. 3 shows that the stability of Rna1p towards α-chymotrypsin (as revealed by the appearance of the 37.5 kDa fragment) is indeed increased provided that the Ca^{2+} concentration exceeds 2–5 mM. Similarly, the generation of the 42.5 kDa fragment is markedly retarded if a limited Asp-N proteolysis is carried out in the presence of 10 mM Ca^{2+} (not shown). As the activity of the two proteases is not significantly affected by Ca^{2+}, the proteolysis patterns indicate that the increased stability results from the binding of one or more Ca^{2+} ions to the Rna1p molecule. It seems likely that the Ca^{2+} binding site(s) reside in the cluster of acidic amino acids present in the C-terminal extensions, as similar clusters of acidic residues have been described to provide low affinity Ca^{2+} sites in other proteins, e.g. calreticulin [22] and calsequestrin [23]. Moreover, the cleavage sites for both proteases, α-chymotrypsin and endoproteinase Asp-N, are located in or close to the highly acidic stretch in the C-terminal extension, and their accessibility is likely to be affected by a direct binding of Ca^{2+} in the close neighbourhood.

4. Discussion

The product of the *RNA1* gene from *S. cerevisiae* has long been implicated in the processing and/or the nucleocytoplasmic transport of different classes of RNA [1,24–26]. Although the molecular basis of the Rna1p action is not known so far, the recent identification of leucine-rich repeats (LRRs) in both *S. cerevisiae* [8] and *S. pombe* [5] Rna1p suggests that Rna1p exerts its function through (a) LRR-mediated protein-protein interaction(s). LRRs, identified in a growing number of

proteins otherwise diverse in function and cellular localisation, have been proposed to represent a versatile binding motif mediating high affinity protein–protein interactions (for review see [16]). This hypothesis is supported by the recently determined crystal structure of the RNase inhibitor, a protein composed entirely of LRRs [15]. The LRRs in this protein form β–α structural units resulting in a non-globular, horseshoe-shaped structure of the molecule in which two principal elements, the α-helices and the β-sheet, are available for interaction with other proteins. The exposed surface of the parallel β-sheet could be of particular importance as such a structure, which usually occurs in the interior of proteins with other structural elements packed against it, could represent the binding site for protein ligands substituting for packing interactions [16].

Within the family of LRR proteins the LRRs present in Rna1p are most closely related to those in the RNase inhibitor, indicating that the structure of the individual β–α units is probably very similar in Rna1p. However, Rna1p contains additional sequence elements, the intervening repeats and the C-terminal extension, which most likely alter the regular structure of the RNase inhibitor. Our proteolysis studies indicate that the C-terminal extension is indeed easily accessible, most likely forming a domain separate from the LRRs and residing on the surface of the molecule. This interpretation is in line with the fact that the majority of the amino acids found in the C-terminal region are the charged residues glutamate or aspartate. Moreover, while LRRs of identical length and number are present in the murine Rna1p homologue, *fug1*, this protein contains an additional 200 amino acids added C-terminally, i.e. has an extended C-terminal domain [6]. Although the functional properties of the C-terminal extension have not been described for the yeast or the murine protein, it is interesting to note that a deletion of this C-terminal region in *S. cerevisiae* Rna1p results in the loss of viability of the respective mutant cells [4].

In contrast to the C-terminal domain the intervening repeats that are inserted after the first and third LRR of all Rna1p species known so far are not particularly exposed to proteolytic attack. Even prolonged treatment of the protein with endoproteinase Asp-N, α-chymotrypsin (Fig. 2), endoproteinase Glu-C, and trypsin (not shown) does not result in a cleavage in or close to the intervening repeats. This indicates that the intervening sequences either form (a) separate but compact structural domain(s) or are embedded in the structure of the LRRs.

The domain structure of Rna1p suggests that the protein could be involved in several functional interactions. One could be mediated through the LRRs, in particular through the non-consensus residues of the repeats. Likewise the intervening repeats could specify the association of certain ligands in conjunction with the neighbouring LRRs. Moreover, the unique C-terminal extension could direct Rna1p to (a) ligand(s) within the cell. Such ligands could be located at the nuclear pore complex as a conditionally lethal mutation in the *nup 1* gene which encodes a nucleoporin in *S. cerevisiae* becomes synthetically lethal in combination with the *rna 1-1* mutation [27].

Acknowledgements: We would like to thank Dr. Klaus Weber for his continuos support and stimulating discussions and Uwe Pleßmann for protein sequence analysis.

References

[1] Hartwell, L. (1967) J. Bacteriol. 93, 1662–1670.
[2] Ruby, S.W. and Abelson, J. (1991) Trends Genet. 7, 79–85.
[3] Osborne, M.A. and Silver, P.A. (1993) Annu. Rev. Biochem. 62, 219–254.
[4] Traglia, H.M., Atkinson, N.S. and Hopper, A.K. (1989) Mol. Cell. Biol. 9, 2989–2999.
[5] Melchior, F., Weber, K. and Gerke, V. (1993) Mol. Biol. Cell 4, 569–581.
[6] DeGregori, J., Russ, A., Von Melchner, H., Rayburn, H., Priyaranjan, P., Jenkins, A.J., Copeland, N.G. and Earl Ruley, H. (1994) Genes Dev. 8, 265–276.
[7] Wilson, R. et al. (1994) Nature 368, 32–38.
[8] Schneider, R. and Schwaiger, M. (1992) Mol. Gen. Genet. 233, 315–318.
[9] Hofsteenge, J., Kieffer B., Matthies, R., Hemmings, B.A. and Stone, S.R. (1988) Biochemistry 27, 8537–8544.
[10] Kataoka, T., Broek, D. and Wigler, M. (1985) Cell 43, 493–505.
[11] Hashimoto, C., Hudson, K.L. and Anderson, K.V. (1988) Cell 52, 269–279.
[12] Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Girandat, J., Leung, J. and Staskawicz, B.J. (1994) Science 265, 1856–1860.
[13] Stone, E.M., Yamano, H., Kinoshita, N. and Yanagida, M. (1993) Curr. Biol. 3, 13–26.
[14] Blackburn, P., Wilson, G. and Moore, S. (1977) J. Biol. Chem. 252, 5904–5910.
[15] Kobe, B. and Deisenhofer, J. (1993) Nature 366, 751–756.
[16] Kobe, B. and Deisenhofer, J. (1994) Trends Biochem. Sci. 19, 415–421.
[17] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.
[18] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[19] Laemmli, U.K. (1979) Nature 227, 680–685.
[20] Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350–4354.
[21] Gerke, V. (1989) FEBS Lett. 258, 259–262.
[22] Smith, M.J. and Koch, G.L.E. (1989) EMBO J. 8, 3581–3586.
[23] Fliegel, L., Oniski, M., Carpenter, M.R., Reithmeier, R.A.F. and MacLennan, D.H. (1987) Proc. Natl. Acad. Sci. USA 84, 1167–1171.
[24] Hutchison, H.T., Hartwell, L.H. and McLaughlin, C.S. (1969) J. Bacteriol. 99, 807–814.
[25] Shiokawa, K. and Pogo, A.O. (1974) Proc. Natl. Acad. Sci. USA 71, 4350–4354.
[26] Hopper, A.K., Banks, F. and Evangelidis, B. (1978) Cell 14, 211–219.
[27] Bogerd, A.M., Hoffmann, J.A., Amberg D.C., Fink, G.R. and Davis, L.I. (1994) J. Cell. Biol. 127, 319–332.