snoRNA are essentially unaffected. Thus, association of UHG RNA with polysomes may enable it to be rapidly degraded; turnover of many cellular mRNAs is known to be coupled to translation[5,6]. Moreover, the densely spaced translational stop codons suggest that the degradation of UHG RNA might be achieved by a nonsense codon-mediated decay mechanism[6].

To verify that the seven conserved UHG intron segments give rise to small RNAs, we analysed HeLa cell nuclear RNA by northern blotting. Figure 3*a* shows that, in addition to intron 9 (which encodes U22), the conserved regions of introns 1 to 4 and 6 to 8 are expressed as different small RNAs, all associated with the nucleolar antigen fibrillarin. Fibrillarin binding correlates with the presence in each of these RNAs of conserved sequences called boxes C and D (Fig. 3*b*) which are found in all metazoan fibrillarin-associated snoRNAs[1]. Several of these UHG-encoded RNAs can also form 4-base-pair (bp) terminal stems commonly observed in this class of snoRNAs (Fig. 3*b*). We designate these novel snoRNAs U25–U31. They range in length from 65 to 76 nucleotides and are similar in abundance to U22 ($\sim 10^4$ molecules per cell). The UHG transcript therefore appears to be a short-lived vehicle for production of the stable U22 (compare lanes 1 and 2 in Fig. 2) and U25–U31 snoRNAs.

U25–U31 all have the potential for extensive base-pairing (12–15 nucleotides) with either 18S or 28S rRNA (Fig. 3*b*). Similar lengthy interactions have been proposed for the vast majority of the intron-encoded snoRNAs[7,8]. Even shorter (6–9 bp) interactions between the non-fibrillarin-associated E1, E2 and E3 snoRNAs and rRNAs have been verified by psoralen crosslinking[9]. The U25–U28/rRNA, U30/rRNA, and U31/rRNA base-pairings are each followed by a CUGA sequence in the snoRNA, or AUGA in the U29 case (Fig. 3*b*). CUGA also appears 3′ to 8 out of 11 other complementarities between fibrillarin-associated snoRNAs and rRNA (our unpublished observation and ref. 8) and therefore can serve as an additional predictor of snoRNA/rRNA interactions. As CUGA is an invariant sequence within box D[1,7], these snoRNAs possess more than one box D-like sequence. We call the second CUGA box D′.

In 18S and 28S rRNAs, the sites complementary to U25–U31 as well as to other snoRNAs, although scattered, fall within the so-called 'universal cores' of the secondary structures[7], implying that these snoRNAs may orchestrate the folding of the pre-rRNA for correct processing and ribosomal protein assembly in the nucleolus. The adjacent CUGA constitutes a conserved RNA structural motif, known as the 'U-turn', in the hammerhead ribozyme[10,11] and transfer RNA[12], and could be used by snoRNAs as a protein-binding signal to facilitate either their interaction or disengagement from rRNA sequences during ribosome biogenesis.

In contrast to most mammalian genes, which discard the bulk of their sequences to produce spliced mRNAs, the eight intron-encoded snoRNAs account for 16 and 20% of the human and mouse UHG transcripts, respectively. Thus, UHG provides a striking example of the utilization of so-called 'junk' DNA, lending support to the existence in eukaryotic organisms of RNA-based gene-expression regulatory systems carried by introns and other apparently non-coding genomic regions[13]. □

Received 18 September; accepted 28 November 1995.

1. Maxwell, E. S. & Fournier, M. J. *A. Rev. Biochem.* **35,** 899–934 (1995).
2. Sollner-Webb, B. *Cell* **75,** 403–405 (1993).
3. Tycowski, K. T., Shu, M.-D. & Steitz, J. A. *Science* **266,** 1558–1561 (1994).
4. Mager, W. H. *Biochem. biophys. Acta* **949,** 1–15 (1988).
5. Greenberg, M. & Belasco, J. G. in *Control of Messenger RNA Stability* (eds Belasco J. & Brawerman G.) 199–218 (Academic, San Diego, 1993).
6. Maquat, L. E. *RNA* **1,** 453–465 (1995).
7. Bachellerie, J.-P. *et al. Trends biochem. Sci.* **20,** 261–264 (1995).
8. Pellizzoni, L., Crosio, C., Campioni, N., Loreni, F. & Pierandrei-Amaldi, P. *Nucleic Acids Res.* **22,** 4607–4613 (1994).
9. Rimoldi, O. J., Raghu, B., Nag, M. K. & Eliceiri, G. L. *Molec. cell. Biol.* **13,** 4382–4390 (1993).
10. Pley, H. W., Flaherty, K. M. & McKay, D. B. *Nature* **372,** 68–74 (1994).
11. Scott, W. G., Finch, J. T. & Klug, A. *Cell* **81,** 991–1002 (1995).
12. Quigley, G. J. & Rich, A. *Science* **194,** 796–806 (1976).
13. Mattick, J. S. *Curr. Opin. genet. Dev.* **4,** 823–831 (1994).
14. Geyer, P. K., Meyuhas, O., Perry, R. P. & Johnson, L. F. *Molec. cell. Biol.* **2,** 685–693 (1982).
15. Nucifora, G., Begy, C. R., Erickson, P., Drabkin, H. A. & Rowley, J. D. *Proc. natn. Acad. Sci. U.S.A.* **90,** 7784–7788 (1993).
16. Tyc, K. & Steitz, J. A. *EMBO J.* **8,** 3113–3119 (1989).

ACKNOWLEDGMENTS. We thank Ch. Cheng for Southern blotting of UHG, S. Huang for the mouse genomic library, S. Wolin for advice on polysome analysis, T. McConnell, E. Scharl, W. Tam and other members of the J.A.S. laboratory for discussion, and J. Mermoud, V. Myer, C. Smith and C. Yoo for critical reading of the manuscript. Supported by a grant from the NIH.

Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry

Matthias Wilm*, Andrej Shevchenko*, Tony Houthaeve*, Stephen Breit†, Lothar Schweigerer†‡, Theodore Fotsis†‡ & Matthias Mann*§

*Protein & Peptide Group, EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany
†Division of Oncology and Haematology, Children's Hospital, University of Heidelberg, INF 150, 69120 Heidelberg, Germany

MOLECULAR analysis of complex biological structures and processes increasingly requires sensitive methods for protein sequencing. Electrospray mass spectrometry[1] has been applied to the high-sensitivity sequencing of short peptides[2], but technical difficulties have prevented similar success with gel-isolated proteins. Here we report a simple and robust technique for the sequencing of proteins isolated by polyacrylamide gel electrophoresis, using nano-electrospray[3,4] tandem mass spectrometry[5,6]. As little as 5 ng protein starting material on Coomassie- or silver-stained gels can be sequenced. Multiple-sequence stretches of up to 16 amino acids are obtained, which identify the protein unambiguously if already present in databases or provide information to clone the corresponding gene. We have applied this method to the sequencing and cloning of a protein which inhibits the proliferation of capillary endothelial cells *in vitro* and thus may have potential antiangiogenic effects on solid tumours.

The peptide extract produced by *in gel* tryptic cleavage of the protein is passed through a capillary which contains a small volume (about 100 nl) of perfusion sorbent (Fig. 1). The adsorbed peptides are washed extensively and step eluted in a volume of approximately 1 μl directly into the electrospray source for mass spectral acquisition. No electroblotting or chromatographic separation is necessary. The procedure is vastly simplified and seems to be more robust than current Edman or mass-spectrometry sample preparation protocols.

The second key component of this method is the nano-electrospray ion source, a miniaturized electrospray source consisting of a metallized glass capillary needle with a tip of inner diameter 1 μm from which the analyte solution is sprayed. Droplets produced by the nano-electrospray are ~ 100 times smaller in volume than those in conventional electrospray sources, allowing efficient use of the whole sample without loss of material in large droplets from which peptides cannot be ionized[3]. The ion current is increased even though the flow rate through the capillary needle

‡ Present addresses: Children's Hospital, University of Marburg, Deutschhausstrasse 12, 35033 Marburg, Germany (L.S.); Laboratory of Biological Chemistry, University of Ioannina Medical School, GR 45110 Ioannina, Greece (T.F.).
§ To whom correspondence should be addressed.

© 1996 Nature Publishing Group

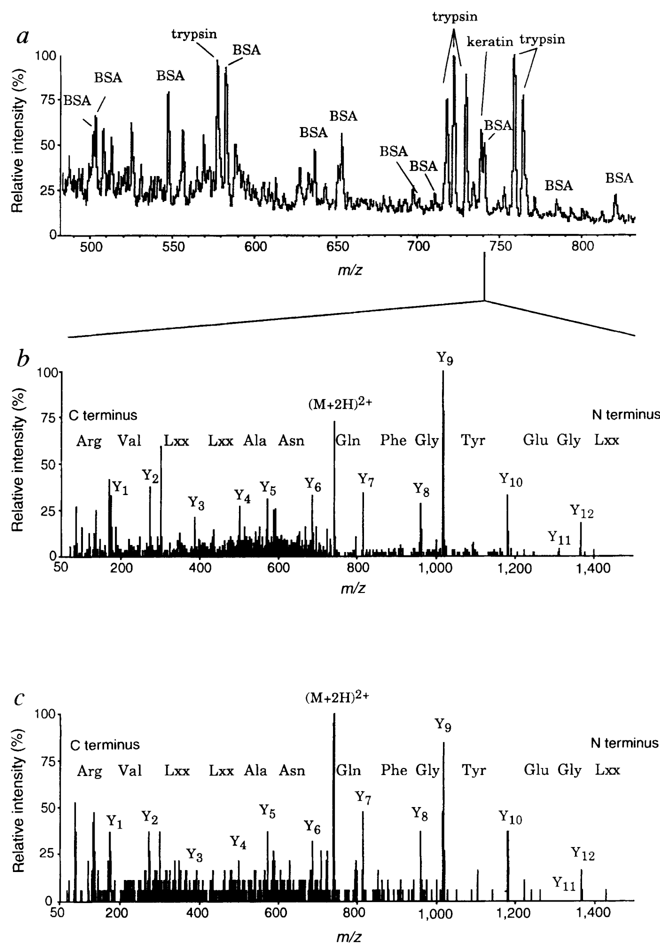


FIG. 1 Testing and sensitivity of the new procedure using BSA. *a*, Peptide ion spectrum of BSA at 800 fmol total material loaded on a gel. Peptide peaks are labelled according to their identification by tandem mass spectrometry. *b*, Fragmentation of the selected doubly charged peptide ion at mass-to-charge *m/z* 740.5. Fragmentation of tryptic peptides at the amide bonds predominantly produces ion series containing the C terminus (designated Y''_1, Y''_2, and so on, see ref. 14). A continuous Y″ ion series[14] could be assigned to the dominant peaks in the spectrum yielding the sequence Lxx-Gly-Glu-Tyr-Gly-Phe-Gln-Asn-Ala-Lxx-Lxx-Val-Arg, where Lxx is either Leu or Ile, which corresponds to a tryptic peptide of BSA. *c*, The same ion as in *b*, obtained in a separate experiment in which 80 fmol BSA was loaded onto a gel that was silver stained.

METHODS. BSA was quantified by amino-acid analysis. Acrylamide gels were prepared using standard protocols and stained with Coomassie blue. *In gel* reduction, acetamidation and tryptic digestion were similar to published procedures[15,16]. After washing with 100 mM NH_4HCO_3 and acetonitrile, gel pieces were swollen in the digestion buffer containing 50 mM NH_4HCO_3, 5 mM $CaCl_2$ and 12.5 ng μl^{-1} trypsin (Boehringer Mannheim, sequencing grade) at 4 °C. After 45 min, the supernatant was aspirated and replaced with 5–10 μl of the same buffer without trypsin to keep gel pieces wet during enzymatic cleavage (37 °C, overnight). Peptides were extracted by three changes of 5% formic acid and acetonitrile and dried down. Approximately 100 nl of POROS R2 sorbent (Perseptive Biosystems) was placed in the tip of a pulled GC 100F-10 (CEI, Pangbourne) capillary. Note that the resin is not packed, and that no frit or other micro LC assembling is necessary. A new capillary and a new portion of resin are used for each analysis to avoid cross-contaminations even at the femtomole level. Dried peptide mixture was dissolved in 10 μl 5% formic acid, loaded onto the pre-equilibrated capillary, washed and eluted with 60% methanol in 5% formic acid into the spraying capillary. The elution volume is 10-fold larger than the resin volume, resulting in good peptide recovery. Nano-electrospray was performed on an API III (Perkin-Elmer Sciex, Ontario, Canada) mass spectrometer as described[4,7]. For precursor ion selection, quadrupole 1 was set to transmit a mass window of 2 Da. Step size for the tandem mass spectra was 0.2 Da, and resolution was set so that fragment masses could be assigned to better than 1 Da.

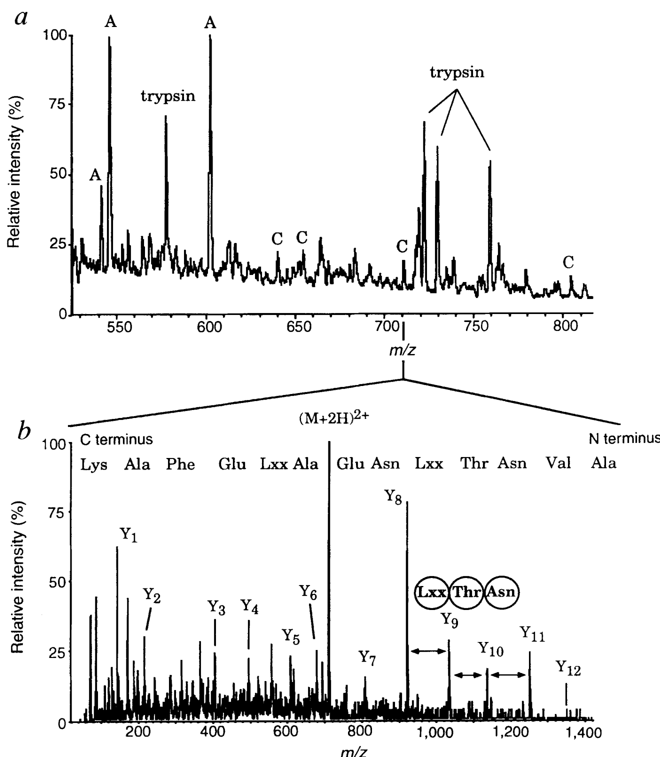


FIG. 2 Identification of protein interaction partners. *a*, Part of the peptide ion spectrum of a tryptic digest of a band of 160K protein excised from polyacrylamide gel. Peptides that were not present in blank were sequenced. The major peaks are tryptic peptides from rabbit antibody light chain (A). *b*, Tandem mass spectrometry of one of the other peptide ions shown in *a*. The masses of the ion and of four fragment ions were assembled by PeptideSearch into a peptide sequence tag[7]. This tag uniquely retrieved carbamoyl-phosphate synthase (M_r 164K) from a non redundant, inclusive sequence database (NRDB, maintained by C. Sander, EMBL). As shown in *b*, the complete series of Y″ ions confirms that the peptide sequence found is correct. Mass spectra of three other ions (labelled C in *a*) identified the same protein.

© 1996 Nature Publishing Group

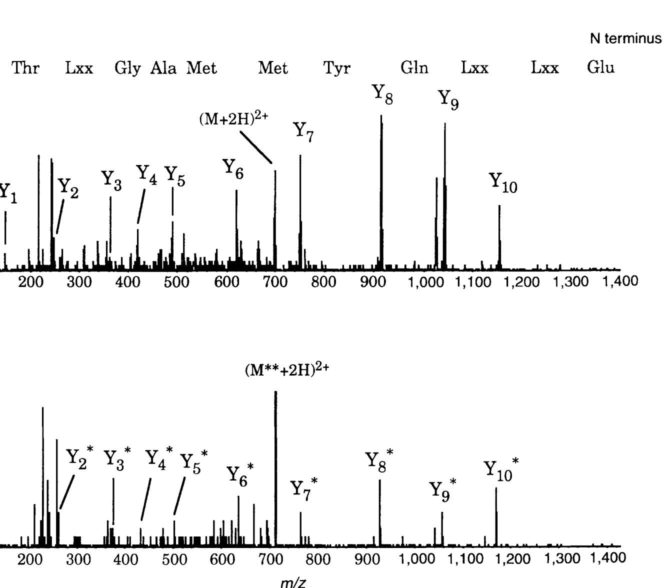


FIG. 3 Sequencing of a peptide from an endothelial cell growth inhibitor. *a*, Tandem mass spectrum of a peptide with m/z 699.3 from the tryptic digest. *b*, Tandem mass spectrum of the same peptide sequenced after esterification[6] of the total digest mixture. Esterification results in a characteristic mass shift of the C terminus containing fragment ions (Y″ ions) by 14 Da and an additional shift of 14 Da for each Asp or Glu residue. The asterisks indicate the incorporated methyl groups. Software-assisted matching of the spectra in *a* and *b* allowed assignment of the sequence with a high degree of certainty.

METHODS. The protein was purified from 20 l conditioned medium from human neurobastoma cells using the inhibition of proliferation of endothelial cells as an assay. After initial concentration using ultrafiltration, the conditioned medium was subjected to cation-exchanged chromatography and chromatofocusing to enrich the inhibitory activity. Final purification was performed using preparative native PAGE followed by preparative isoelectric focusing in tubes (T.F., in preparation). SDS–PAGE analysis indicated the presence of a protein of M_r 45K (both under reducing and non-reducing conditions) and a yield of approximately 400 ng. Half of the gel piece containing 200 ng of the purified protein was subjected to the new mass-spectrometric sequencing procedure.

is extremely low (20–40 nl min^{-1}). Very small amounts (1–2 μl) of the peptide mixture can be subjected to mass spectrometric analysis over 30–120 min. Thus there is sufficient measurement time to optimize tandem mass spectrometric conditions for each of the peptide ions in turn. Tuning the collision energy individually was found to be important to obtain the best possible tandem mass spectra for peptide sequence determination.

We tested the new procedure on a sample of bovine serum albumin (BSA). Figure 1*a* shows the initial peptide ion spectrum, and Fig. 1*b* a tandem mass spectrum, resulting from loading a total of 800 fmol (53 ng) of the protein on a gel. In addition to trypsin and ubiquitous human keratin contaminations, 11 BSA peptides were sequenced. Unambiguous sequence information was also obtained from four peptides when only 80 fmol (5 ng) of BSA was loaded onto the gel (Fig. 1*c*). In the latter case, the protein band was silver stained because it was hardly visible after Coomassie staining. To the best of our knowledge, there have been no previous reports of protein sequencing from silver-stained gels.

The developed methodology was applied to the identification of protein interaction partners. Approximately 100 ng of an immunoaffinity-purified protein interacting with an RNA binding protein was isolated by gel electrophoresis. The mass spectrum of the peptides derived from this band (Fig. 2*a*) showed high-intensity peaks, which were quickly assigned to trypsin and to antibody (used in the purification step) through database searches by peptide sequence tags[7]. Four other peptides (from Fig. 2*a*) were found to belong to a known protein in the database (Fig. 2*b*). The small amount of sample, the excess of antibody over

TABLE 1 Comparison of the deduced amino-acid sequence to the partial sequences obtained by tandem mass spectrometry

Nucleic acid sequence of cloned gene	Deduced amino-acid sequence	Amino acid sequence obtained by mass spectrometry
GAA CCT GTT CTT TCA CCT GAA CAC AGA	Glu-Pro-Val-Leu-Ser-Pro-Glu-His-Arg	Glu-Pro-Val-Lxx-Ser-Pro-Glu-His-Arg
GAA TTA ATC CAA TAC ATG ATG GCA GGT ATT ACT AAA	Glu-Leu-Ile-Gln-Tyr-Met-Met-Ala-Gly-Ile-Thr-Lys	Glu-Lxx-Lxx-Gln-Tyr-Met-Met-Ala-Gly-Lxx-Thr-Lys
GAC CCA TTT GCA TCA GTT GGT AAT GGT GTT ACA ATT CAC TAC ATG CGT	Asp-Pro-Phe-Ala-Ser-Val-Gly-Asn-Gly-Val-Thr-Ile-His-Tyr-Met-Arg	Asp-Pro-Phe-Ala-Ser-Val-Gly-Asp-Gly-Val-Thr-Lxx-His-Tyr-Met-Arg
ACT CCA TGA TAC TAT GAC CCA GCA ATG AAA	Thr-Pro-Trp-Tyr-Tyr-Asp-Pro-Ala-Met-Lys	Thr-Pro-Trp-Tyr-Tyr-Asp-Pro-Ala-Met-Lys
ACT GAC CTA GAA ACT ATT ACT TTA TTA GCT AAA	Thr-Asp-Leu-Glu-Thr-Ile-Thr-Leu-Leu-Ala-Lys	Thr-Asp-Lxx-Glu-Thr-Lxx-Thr-Lxx-Lxx-Ala-Lys
ATA GTT GCA ATT AAT GTT CCT AAA	Ile-Val-Ala-Ile-Asn-Val-Pro-Lys	Lxx-Val-Ala-Lxx-Asn-Val-Pro-Lys
TGT ATG TCA ATG CCT CTT TCA CGT	Cys-Met-Ser-Met-Pro-Leu-Ser-Arg	Cys-Met-Ser-Met-Pro-Lxx-Ser-Arg

The peptide Cys-Met-Ser-Met-Pro-Lxx-Ser-Arg (Lxx is Leu or Ile), which was at the most C-terminal position (as judged by homology), was used to generate an antisense PCR primer (5′-TGATAAAGGCATTGACATAC-3′). As sense primer, a sequence from the 5′-conserved region of the *Mycoplasma arginini* (5′-AGGAATTCACGTTTATTCAG-3′) was used to obtain a nearly full-length clone using DNA prepared from the contaminated tumour cells as template. The sequence has been deposited with the EMBL data library (accession no. X93471). The only discrepancy observed was the prediction of Asn from the nucleotide sequence of the cloned protein instead of the mass spectrometrically sequenced Asp in one of the peptides (underlined). Because the identity of Asp is absolutely certain owing to the +14 Da mass shift induced on this amino acid by the derivatization procedure in conjunction with the mass-spectrometric data on the underivatized peptide, it is concluded that deamidation of the amino acid must have occurred during the purification procedure, an event not unusual during protein purification. Thus the amino-acid sequence provided by mass spectrometry was entirely correct.

© 1996 Nature Publishing Group

the protein of interest, and its large size (M_r 164K) would have made identification by conventional techniques exceedingly difficult.

Nano-electrospray mass spectrometry was further used in our search for angiogenesis inhibitors, that is, factors that suppress the formation of new blood vessels[8,9]. We observed that conditioned media from tumour-cell cultures contaminated by mycoplasma contained a protein that strongly inhibited the *in vitro* proliferation of endothelial cells, the key cell involved in angiogenesis. SDS–polyacylamide gel electrophoresis (PAGE) analysis after purification revealed a protein with a M_r of 45K. About 200 ng of the protein was subjected to this procedure (Fig. 3). Seven peptides were sequenced in their native and esterified forms, and 73 amino acids were called. A database search of these peptides indicated a high degree of homology to arginine deiminase from *Mycoplasma arginini*. By using the obtained sequence and the homology, we have generated oligonucleotide primers and cloned the complementary DNA by polymerase chain reaction (PCR). A comparison of the deduced amino-acid sequence to the partial sequences obtained by tandem mass spectrometry is made in Table 1, which shows that all 73 amino acids were sequenced correctly. It is clear from the data that the accuracy and the amount of the amino-acid sequence data would have made cloning by PCR[10] feasible even without homology. Experiments are underway to determine the role of this protein in the inhibition of angiogenesis and its potential in the antiangiogenic treatment of solid tumours.

In conclusion, amino-acid sequence data can now be obtained by mass spectrometry with extremely small quantities of gel-isolated proteins. This procedure is ∽10–100 times more sensitive than current techniques, which have a practical limit of 10–50 pmol loaded on the gel for protein microsequencing based on *in gel* or on-blot digestion followed by Edman degradation[11–13]. It is also much faster and involves fewer purification and manipulation steps. We anticipate that sequencing by nano-electrospray tandem mass spectrometry will thus become a method of choice for protein microcharacterization. In conjunction with genomic sequencing, it will allow the rapid functional characterization of proteins with important biological functions. □

Received 27 September; accepted 28 November 1995.

1. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. *Science* **246,** 64–71 (1989).
2. Hunt, D. F. *et al. Science* **255,** 1261–1263 (1992).
3. Wilm, M. S. & Mann, M. *Int. J. Mass Spectrom. Ion Process.* **136,** 167–180 (1994).
4. Wilm, M. & Mann, M. *Analyt. Chem.* **68,** 1–8 (1996).
5. Biemann, K. & Scoble, H. A. *Science* **237,** 992–998 (1987).
6. Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S. & Hauer, C. R. *Proc. natn. Acad. Sci. U.S.A.* **83,** 6233–6237 (1986).
7. Mann, M. & Wilm, M. S. *Analyt. Chem.* **66,** 4390–4399 (1994).
8. Fotsis, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **90,** 2690–2694 (1993).
9. Fotsis, T. *et al. Nature* **368,** 237–239 (1994).
10. Walker, J. E. *et al. J. molec. Biol.* **226,** 1051–1072 (1992).
11. Hunkapiller, M. *et al. Nature* **310,** 105–111 (1984).
12. Erdjument-Bromage, H., Lui, M., Sabatini, D. M., Snyder, S. H. & Tempst, P. *Protein Sci.* **3,** 2435–2446 (1994).
13. Matsudaira, P. (ed.) *A Practical Guide to Protein and Peptide Purification for Microsequencing* (Academic, San Diego, CA, 1993).
14. Roepstorff, P. & Fohlman, J. *Biomed. Mass Spectrom.* **11,** 601 (1984).
15. Jenö, P., Mini, T., Hintermann, E. & Horst, M. *Analyt. Biochem.* **224,** 451–455 (1995).
16. Rosenfeld, J., Capdevielle, J., Guillemot, J. C. & Ferrara, P. *Analyt. Biochem.* **203,** 173–179 (1992).

ACKNOWLEDGEMENTS. M.W. and A.S. contributed equally to this work. We thank T. Hyman, E. Izaurralde and A. I. Lamond for cooperation and discussion, and F. Kafatos for criticism and help with the manuscript. This work was supported by the BMBF and constitutes partial fulfillment of the requirement for the Ph.D. of M. Wilm. Work at the University of Heidelberg is supported by grants from Deutsche Forschungsgemeinschaft and Deutsche Krebshilfe.

KNOW YOUR COPY RIGHTS

RESPECT OURS

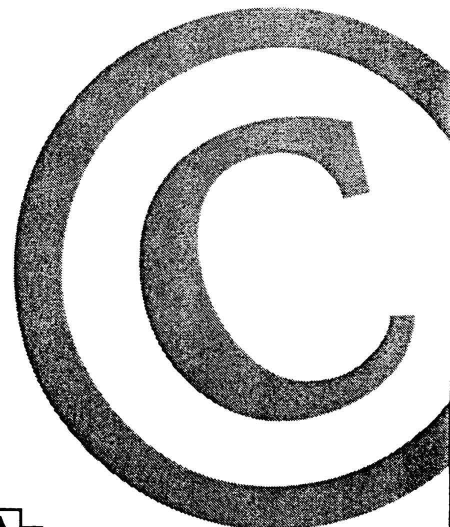
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP

Telephone: 0171 436 5931
Fax: 0171 436 3986