Rapid identification of DNA-binding proteins by mass spectrometry

Eckhard Nordhoff[1,3*†], Anne-M. Krogsdam[1†], Helle F. Jørgensen[2], Birgitte H. Kallipolitis[1], Brian F.C. Clark[2], Peter Roepstorff[1], and Karsten Kristiansen[1]

[1]Department of Molecular Biology, Odense University, Odense, Denmark. [2]Department of Biostructural Biology, Institute of Molecular and Structural Biology, University of Aarhus, Aarhus, Denmark. [3]Current address: Max Planck Institute for Molecular Genetics, Berlin, Germany.
**Corresponding author (e-mail: nordhoff@mpimg-berlin-dahlem.mpg.de).†These authors contributed equally to this work.*

Received 12 November 1998; accepted 26 May 1999

We report a protocol for the rapid identification of DNA-binding proteins. Immobilized DNA probes harboring a specific sequence motif are incubated with cell or nuclear extract. Proteins are analyzed directly off the solid support by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. The determined molecular masses are often sufficient for identification. If not, the proteins are subjected to mass spectrometric peptide mapping followed by database searches. Apart from protein identification, the protocol also yields information on posttranslational modifications. The protocol was validated by the identification of known prokaryotic and eukaryotic DNA-binding proteins, and its use provided evidence that poly(ADP-ribose) polymerase exhibits DNA sequence-specific binding to DNA.

Keywords: protein–DNA interactions, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

Although a substantial number of proteins involved in transcriptional regulation have been identified, a majority are likely still unknown. Current genetic strategies such as the one-hybrid assay[1] and phage-display techniques[2] suffer from the inability to detect proteins whose specific binding to a DNA element is dependent upon accessory proteins. In this paper we describe an approach that relies on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)[3] to identify DNA-binding proteins isolated from cell extracts by virtue of their interaction with double-stranded DNA probes immobilized onto small, paramagnetic particles. Bound proteins are directly analyzed by MALDI-TOF MS, in which formed molecular ions are separated and detected according to their mass-to-charge (m/z) ratio. Singly charged ions predominate in MALDI-TOF MS, making interpretation of spectra of multicomponent samples straightforward. The analysis requires only femtomole amounts of proteins and allows determination of molecular masses with deviations of at most a few Daltons in less than 30 min.

In cases where determination of molecular masses does not permit unambiguous identification of proteins, the samples can be passed over an immobilized-trypsin microcolumn[4] to generate a mixture of peptides from the individual proteins. The molecular masses of the peptides are then determined by MALDI-TOF MS and compared with those predicted from virtual tryptic digests of protein sequences stored in databases. The procedure requires only low-femtomole amounts of analytes and yields mass determinations deviating less than 40 p.p.m. The identified proteins are further characterized with respect to possible posttranslational modifications by comparing the protein molecular weights and peptide map data with the retrieved sequences. In cases in which the protein mixture is too complex for direct peptide mapping, gel electrophoresis followed by in situ digestion and mass spectrometric peptide mapping can be applied[5]. This extension requires more time and no less than a few nanograms of each protein but allows the unambiguous identification of many proteins simultaneously.

Results and discussion

Identification of cyclic adenosine monophosphate receptor protein. An immobilized double-stranded oligodeoxynucleotide probe harboring a binding site for cyclic adenosine monophosphate (cAMP) receptor protein (CRP)[6–7] was used to isolate proteins from crude cell extracts of Escherichia coli strain SØ928 (ref. 8). Spectra were obtained from a 10 µg fraction from 100 µg beads incubated with ~2 µg total protein (Fig. 1B). cAMP was added to a concentration of 100 µM to ensure high abundance of the CRP–(cAMP) complex[9]. We detected two dominant signals at m/z 23,500 and m/z 11,750 that were absent from samples probed with either a control oligonucleotide, lacking a DNA probe, or from unprocessed cell extract (Fig. 1 C, D, and A, respectively). These signals represent the singly and doubly protonated CRP molecular ions (determined molecular mass: 23,510.6; expected: 23,509.4). In the absence of cAMP, the amount of bound CRP was dramatically reduced (data not shown). A weak signal at m/z 13,000 was present in all three spectra and corresponded to streptavidin monomers extracted from the Dynabeads (Dynal, Oslo, Norway) providing an internal reference for spectra calibration.

MALDI-TOF MS peptide mapping. For unambiguous identification, proteins bound to a 20 mg portion of the CRP consensus beads were eluted and digested on an immobilized-trypsin microcolumn followed by MALDI-TOF MS analysis (Fig. 1E). Within an allowed deviation of 40 p.p.m., 13 out of 25 selected masses matched tryptic peptides of CRP, covering 72% of the sequence altogether. In contrast, colicin E1 protein matched only five peptides. In a second run, the NCBI (release April 7, 1999) and the Swissprot (release April 7, 1999) databases were searched with no species restriction. Again, *E. coli* CRP was identified as the only likely candidate followed by CRP from *Salmonella typhimurium*. The molecular mass calculated from the retrieved sequence differs from the experimentally determined value by 130.0 Da (Fig. 1E) indicating removal of the N-terminal methionine residue (expected mass difference: 131.2 Da). Using the shortened sequence, the detected monoisotopic peptide

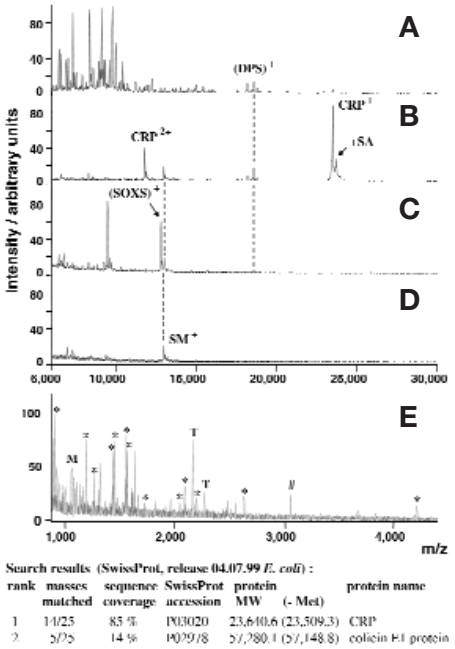


Figure 1. Identification of cyclic adenosine monophosphate receptor protein (CRP) in E. coli crude cell extract. (A) Mass spectrum obtained from 0.01 μl cell extract (input). Mass spectra of proteins isolated with an immobilized DNA probe containing the CRP-binding sequence (B), lacking the CRP-binding sequence (C), or lacking DNA (D). +SA, Adduction of one sinapic acid molecule used as MALDI matrix. SM, Streptavidin monomer. (E) Identification of CRP by peptide mapping. *Identified CRP tryptic peptides; #the first 26 amino acid residues of CRP lacking the N-terminal methionine. M, Matrix signal. T, Trypsin autolysis products, used for internal spectrum calibration. DPS, DNA-binding protein from starved cells.

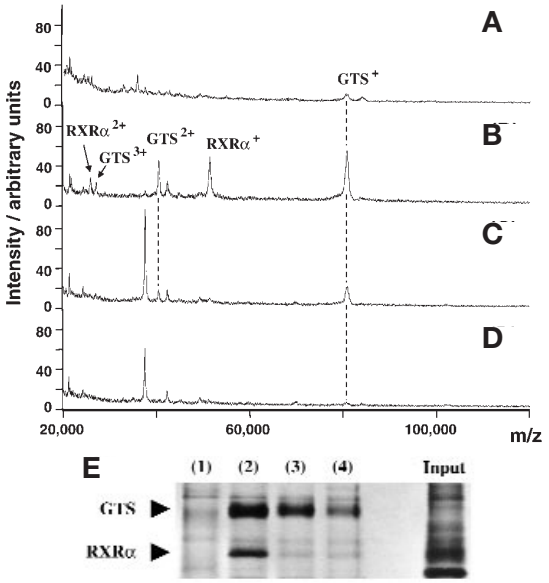


Figure 2. Identification of ectopically expressed rat RXRα in cell extracts from control yeast (A) or yeast expressing rat RXRα (B). (C and D) As in (B) but using two different immobilized DNA probes carrying no DR1 sequence element. GTS, glutamyl-tRNA synthetase signals. (E) SDS–PAGE analysis of proteins from (A–D), respectively. "Input" represents 1 μl extract of cells expressing rat RXRα. The section of the gel covering the molecular mass range 50,000–100,000 Da is shown. The positions of RXRα and GTS are indicated.

mass 3,061.55 Da (Fig. 1E) could be assigned to the singly protonated peptide comprising the first 26 residues. This result confirmed removal of the initiating methionine and increased the sequence coverage from 72% to 85%.

A signal at m/z 18,550 was detected both with the consensus CRP oligonucleotide probe and the control probe (Fig. 1B and C), indicating non-specific binding. Allowing a deviation of ±2.5 Da for the determined molecular mass, DNA-binding protein from starved cells was identified as a likely candidate from the databases listed above (expected/determined molecular mass: 18,564.2 Da/18,562.6 Da). An intense signal at m/z 12,800 present in the control spectrum (Fig. 1C) yielded the transcriptional activator regulatory protein SOXS (12,779.7 Da/12,780.4 Da) as a likely candidate. The signal at m/z 9,450 (9,452.5 Da) did not match any database entry.

Identification of rat retinoid X receptor and yeast glutamyl-tRNA synthetase. Rat retinoid X receptor-α (RXRα) is a member of the nuclear hormone receptor family of higher eukaryotes. We expressed rat RXRα in yeast strain BJ2168[10] and used a DNA probe that contained the DR1 element from the rat acyl-CoA oxidase promoter. This element has been shown to direct RXRα-dependent transactivation of a reporter gene in yeast[11]. The MALDI-TOF MS spectra of proteins isolated from crude extracts of cells expressing RXRα was compared with spectra obtained from control cell extracts and with spectra from RXRα-expressing cells obtained using probes lacking the DR1 motif (Fig. 2). Two abundant signal series comprising singly, doubly, and triply charged molecular ions were observed in spectra obtained from proteins isolated from RXRα-expressing cells using the RD1 probe. The first signal series represents RXRα and is not observed in the other spectra. The protein in the second series was identified as yeast glutamyl-tRNA synthetase (GTS) and is present in all four spectra. The proteins pulled down by the DR1 probe were separated by SDS–PAGE, and the protein bands shown in Figure 2E were subjected to in situ tryptic digestion followed by mass spectrometric peptide mapping and database search.

The molecular mass determined for RXRα (51,147 Da) deviates by 215 p.p.m. from the expected value (51,136 Da), a reasonable result for this mass range. In the case of GTS, however, the deviation of more than 4,000 p.p.m. (observed, 80,875 Da; expected, 82,533 Da) indicates substantial truncation of GTS. Whether this processing is of biological significance remains to be established. Yet, the examples demonstrate that different proteins binding to the same target sequence can be identified in a single run and that the determined molecular masses provide valuable information.

Recently, the phenylalanyl-tRNA synthetase from *Thermus thermophilus* was demonstrated to bind specifically to genomic DNA sequences[12]. Intriguingly, binding of GTS to the DR1 probe was significantly enhanced in the presence of RXRα (Fig. 2A and B). Lower levels of GTS also bound the control DNA probes, but both SDS–PAGE and MALDI-TOF MS analysis indicated that the highest level of binding of GTS occurred using the DR1 probe on extracts of cells expressing RXRα (Fig. 2B and E). The biological significance of this observation is unclear, but the recent report that RXRα specifically interacts with the T:G mismatch-specific thymidine-DNA glycosylase[13] suggests that functional RXRα interactions may extend beyond proteins involved in transcriptional regulation.

Expression of RXRα and the nuclear hormone receptor peroxisome proliferator-activated receptor γ (PPARγ) are enhanced upon induction of differentiation of the mouse adipocyte cell line, 3T3-F442A, by treatment with insulin and triiodothyronine. The heterodimeric complex PPARγ–RXRα has been shown to function as a key regulator of adipocyte differentiation[14]. To test the utility of the technique for identification of endogenous RXRα in a mammalian extract, we isolated proteins from crude nuclear extract of induced 3T3-F442A cells using the DR1 oligonucleotide probe. Analysis by SDS–PAGE identified a 55 kDa protein specifically captured by the DR1 probe (Fig. 3). Mass spectrometric peptide mapping and data-

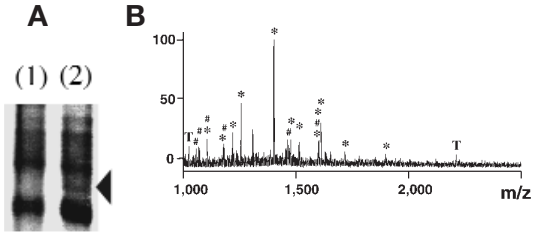


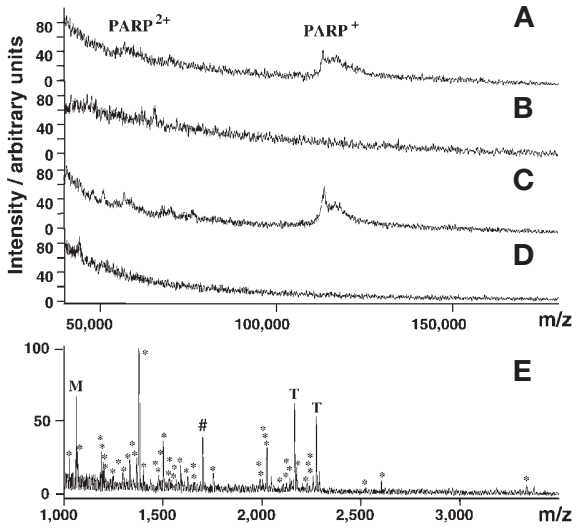
Figure 3. Identification of RXRα and PPARγ in 3T3-F442A crude nuclear extract. (A) Section of the gel containing proteins retained with the immobilized control (1) or DR1 (2) probe. (B) Tryptic peptide map of the proteins contained in the marked band. *Identified digestion products of RXRα; #identified digestion products of PPARγ2. T, Trypsin autolysis products used for internal spectrum calibration.

base search identified mRXRα as the most likely candidate (12 hits out of 25 selected masses) followed by a second nuclear receptor, mPPARγ (six hits). In a second search, performed with the peptides that were not identified as tryptic cleavage products of RXRα, mPPARγ was identified as the most likely candidate. This identification was verified by western blot analysis (data not shown).

Identification of poly(ADP-ribose) polymerase. We applied our protocol to identify a protein binding to an element termed EFP1 involved in regulation of expression of the elongation factor-1 α gene[15]. Electrophoretic mobility shift assays (EMSAs) performed with nuclear extracts of HeLa cells revealed specific binding of nuclear protein(s) to the EFP1 probe. Strong binding was observed even in the presence of 500 mM KCl (Fig. 4). This protein was identified as poly(ADP-ribose) polymerase (PARP), following the strategy described above (Fig. 5). The assigned peptides covered 55% of the human PARP sequence, followed by mouse PARP with only 33% sequence coverage. This result identified human PARP as the only likely candidate. The determined peptide masses provided strong evidence that the N terminus of PARP is acetylated. An intense signal at m/z 1,700 matched the mass expected for a singly acetylated tryptic fragment comprising the first 14 residues (expected/measured monoisotopic mass: 1699.84 Da/1699.89 Da) and did not match any other tryptic peptide of PARP. The acetylation increased the expected molecular mass from 112,952 Da to 112,994 Da, thereby reducing the deviation from the experimentally determined value (113,057 Da) to 560 p.p.m.; a reasonable value for molecular masses exceeding 100,000 Da.

Poly(ADP-ribose) polymerase is a chromatin-associated protein that modifies various nuclear proteins by poly(ADP-ribosyl)ation in a DNA-dependent manner[16]. It is thought to take part in regulation of differentiation, proliferation, tumor transformation, and DNA repair, and proteolytic cleavage of PARP is a marker for the onset of apoptosis[17]. Recently, PARP was reported to participate in transcriptional activation[18]. Here, we provide the first evidence that PARP recognizes a specific DNA sequence in vitro. The immobilized DNA probe did not contain single-stranded overhangs, and an excess of the nonbiotinylated oligonucleotide was used in the annealing procedure to minimize the inclusion of single-stranded DNA during isolation, since PARP is known to bind single-stranded DNA breaks. Control experiments using the biotinylated DNA strand alone showed no PARP binding (results not shown). A possible role for PARP in transcriptional regulation of elongation factor-1 α is being investigated. Besides the signal assigned to PARP, the protein spectrum shows additional overlapping peaks spanning the range m/z 114,000–122,000. These signals are indicative of partial auto-poly(ADP-ribosyl)ation, a well-known feature of PARP[16]. According to our data, PARP molecules modified with up to 18 ADP-ribose units were able to bind to the EFP1 element. The unmodified protein is the most abundant species followed by species harboring six to seven attached ADP-ribose units (second maximum at m/z 116,500).

Major obstacles in the identification of sequence-specific DNA binding proteins relate to large variation in abundance and specific versus nonspecific DNA binding affinities. False positives are relatively easy to eliminate by the use of control DNA probes. False negatives due to suboptimal binding and washing conditions are more difficult to circumvent and may require more elaborate adjustments of experimental conditions. In terms of sensitivity the MALDI-TOF MS protocol for analysis of DNA–protein interaction is still inferior



Search results (SwissProt, release 04.07.99, all species) :

rank	masses matched	sequence coverage	SwissProt accession	species	protein MW	protein name
1	45/56	55 %	P09874	human	112,953	PARP
2	24/56	33 %	P11103	mouse	112,969	PARP

Figure 5. Identification of poly(ADP-ribose) polymerase (PARP) captured by the EFP1 probe. A 400 μg sample of Dynabeads was incubated with 40 ml HeLa crude nuclear extract in a final volume of (A and B) 550 μl containing 50 mM KCl, and (C and D) 55 μl containing 500 mM KCl. Analysis by MALDI-TOF MS was carried out on 20 μg Dynabeads pull-down portions. (A and C) Mass spectra obtained with immobilized DNA containing the EFP1 element. (B and D) Mass spectra obtained with an unrelated DNA sequence. Comparison of (A) and (C) demonstrates that more PARP is captured in a low-volume high-salt binding reaction (C) than in a large-volume low-salt binding reaction (A). (E) Identification of PARP by peptide mapping. Proteins pulled down with 300 μg EFP1-Dynabeads under low-salt conditions were separated by SDS–PAGE followed by in situ tryptic digestion. The spectrum was acquired from 0.5 μl of a total of 20 μl supernatant obtained from the digestion of the PARP band. *Identified digestion products of PARP; #singly acetylated tryptic peptide 1–14. M, Matrix signal; T, trypsin autolysis products, used for internal spectrum calibration.

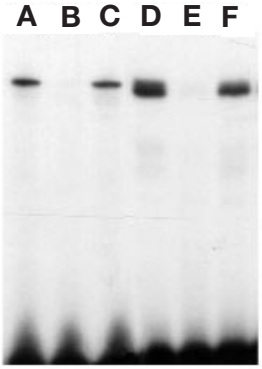


Figure 4. Specific binding of HeLa cell nuclear proteins to the EFP1 element. Binding to the EFP1 element was analyzed by standard EMSA. HeLa crude nuclear extracts were incubated with the labeled EFP1 probe (probe 5; see Experimental Protocol) in low (A–C) and high (D–F) -salt binding buffer. (A and D) No competition. (B and E) Competition with a 20-fold excess of cold EFP1 probe. (C and F) Competition with a 20-fold excess of an irrelevant probe (probe 7).

to EMSA technology, which stands out as the obvious supplementary protocol for confirmation of results. It is well known that MALDI-TOF MS, at the current state of development, does not provide quantitative data because of sample heterogeneity and the observation that signal intensities depend on the primary structure of the analyte molecules as well as sample complexity in case of mixtures. Because of this restriction, our protocol does not include determination of binding constants or binding stoichiometries.

The solid support is well suited for high-throughput applications, and the necessary instrumentation is commercially available. The use of paramagnetic particles is especially efficient for large-scale screenings in which numerous samples can be examined in parallel using different binding and washing conditions. Such samples may comprise extracts of cells from different organisms, different developmental stages, as well as extracts of cells challenged with different drugs. In many of these experiments, peptide mapping will not be required. However, the power of MALDI-TOF MS peptide mapping for identification of posttranslational modifications and the possibility of analyzing multiprotein complexes add a new dimension to such experiments. The method should also be applicable to proteins that bind specifically to single-stranded DNA or RNA. Finally, combined with DNA chip technology, the protocol allows for efficient combinatorial screening applications, and this approach would also pave the way for systematic analysis of drug-induced changes in DNA–protein interactions.

Experimental protocol

DNA probes: (1) 5'-biotin-d(TTCAAGGCGAAAAGTGTGACATATGTCACACTTTTCGCCTCTT)-3', 5'-(TAAGAGGCGAAAAGTGTGACATATGTCACACTTTTCGCCTTGA)-3'; (2) 5'-biotin-d(TTCCGCAAACAAAACCCAAAAACAACCCATACCCAACCCTCGTGC)-3', 5'-d(GCAAGAGGGTTGGGTATGGGTTGTTTTTGGGTTTTGTTTGCGGAA)-3'; (3) 5'-biotin-d(TCGACTCCCGAACGTGACCTTTGTCCTGGTCCCCTGTCGAC)-3', 5'-d(GTCGACAGGGGACCAGGACAAAGGTCACGTTCGGGAGTCGA)-3'; (4) 5'-biotin-d(CCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAG)-3', 5'-d(CTTTCCCAGTTTACCCCGCGCCACCTTCTCTAGG)-3'; (5) 5'-d(AGCTTGGCGCGGGGTAAACTGGGAAAGTCG)-3', 5'-d(TCGACGACTTTCCCAGTTTACCCCGCGCCA)-3'; (6) 5'-biotin-d(AGTTGCCTAGGCATATTGCGCAATATGC)-3', 5'-d(GCATATTGCGCAATATGCCTAGGCAAC)-3'; (7) 5'-d(CTAGATGTACTGGCTCCGCCTTTTTCCCGG)-3', 5'-d(GATCCCGGGAAAAAGGCGGAGCCAGTACAT)-3'

Preparation of cell extracts. *Escherichia coli* (strain SØ928 [ref. 8]) crude cell extract was prepared as described[19]. *Saccharomyces cerevisiae* strain BJ2168 was transformed with plasmid pRS-315 containing or lacking the rRXRa cDNA under control of the yeast CUP1 promoter. Cultures were grown overnight in selective media containing 0.2 mM CuSO4. For preparation of whole-cell extracts the cells were harvested, resuspended in two volumes buffer (0.4 M KCl, 20 mM Tris-HCl, pH 7.5, 20% glycerol, 2 mM dithioerythritol (DTE), 2 mM $MgCl_2$ and protease inhibitors) and disrupted by one freeze-thaw cycle and five passes through a French press (1,300 Psi). The lysates were centrifuged at 20,000 r.p.m. (Beckman JA20 rotor; Beckman Instruments, Fullerton, CA). The 3T3-F442A cells were grown to confluence in Dulbecco's modified Eagles medium (DMEM) supplemented with 8% bovine serum. At confluence, 10 nM insulin and 2 nM triiodothyronine were added to the medium to induce adipocyte differentiation. Eight days post-confluence the cells were harvested and crude nuclear extracts were prepared[20]. Nuclear extracts of HeLa cells were prepared as described[21]. EMSAs were performed with 3 µg of crude HeLa nuclear extract using a ^{32}P-labeled EFP1 DNA probe (probe 5) with or without unlabeled competitor probes (probes 5 or 7) as described[20].

Dynabead purifications. Approximately 0.2 pmol of double-stranded DNA was immobilized onto 1 µg Dynabeads M-280 Streptavidin (Dynal) following the protocol supplied by the manufacturer. For identification of CRP, 0.05 µl crude *E. coli* cell extract in a total volume of 400 µl containing 100 mM NaCl, 20 µM Tris-HCl, pH 7.5, 1 µM DTE, 100 µM cAMP, protease inhibitors (Complete; Boehringer Mannheim, Indianapolis, IN), and 0.1 µg/µl poly-d(IC) were preincubated on ice for 10 min before addition of 100 mg Dynabeads with immobilized DNA probes. The mixture was incubated for 15 min at room temperature in an end-over at 20 r.p.m.. After magnetic separation, the beads were washed three times in 80 µM NaCl, 20 µM Tris-HCl, pH 7.5, 1 µM DTE, and protease inhibitors. The first wash also contained 0.1 µg/µl poly-d(IC). Thereafter, the beads were washed once in 20 µl of 20 mM Tris-HCl, pH 7.5, 1 mM DTE, and subsequently rinsed with 10 µl of the same buffer. A 10 µg portion of Dynabeads from each sample was subjected to MALDI-TOF MS analysis.

For identification of RXRα, 200 mg Dynabeads with immobilized DNA probes and 80 µl yeast cell extract were used according to the procedure described above, except that the binding reaction additionally contained 10% glycerol, 4 mM $MgCl_2$, and as competitor DNA 1 µg poly-d(IC) plus 1 mg sonicated herring sperm DNA per 10 µg protein were included. The first washing buffer was further supplemented with 40 µg/µl sonicated herring sperm DNA. Subsequently, 20 mg portions of each Dynabeads isolate were analyzed by MALDI-TOF MS. Then 100 mg portions were resuspended in 30 ml Laemmli sample buffer containing 7 M urea and separated on a 10% SDS–polyacrylamide gel.

The same pull-down protocol was applied to 50 µl crude 3T3-F442A nuclear extract using 400 µg Dynabeads with immobilized DNA probe (probe 3 or 6). For pull-downs from crude HeLa cell nuclear extracts, Dynabeads carrying the DNA probe (probe 3 or 4) were used. Two different binding reactions were performed. In one reaction, 40 µl crude nuclear extract (~150 µg protein) were diluted to a final volume of 550 µl, reducing the KCl concentration to 50 mM (large-volume low-salt binding). This solution also contained 10% glycerol, 20 mM HEPES, pH 7.9, 1 mM $MgCl_2$, 1 mM DTE, 0.1 µg/µl poly-d(IC), and protease inhibitors. In a second run, the reaction volume was reduced to 55 µl, corresponding to 500 mM KCl (low-volume high-salt binding). In both cases, the reaction mixtures were preincubated for 10 min at room temperature before 400 µg Dynabeads were added, and then incubated for another 25 min. After magnetic separation, the beads were washed three times with 400 ml buffer (50 mM KCl, 20 mM HEPES, pH 7.9, 1 mM $MgCl_2$, 0.5 mM DTE, protease inhibitors). The first washing included 0.1 µg/µl poly-d(IC). The captured proteins were analyzed as described above except that a 300 µg Dynabeads aliquot was loaded onto a 7% SDS–polyacrylamide gel.

MALDI-TOF MS analysis. Dynabeads from pull-downs were resuspended in 0.3 µl 20 mM Tris-HCl and transferred onto the MALDI sample support followed by addition of 0.3 µl saturated solution of sinapic acid in 35% acetonitrile, 1% TFA (matrix solution). After solvent evaporation, the sample was loaded into a Voyager Elite MALDI-TOF mass spectrometer (PerSeptive, Cambridge, MA). Typically, 50–100 single-shot spectra were averaged for improved signal-to-noise ratio. For unambiguous identification, proteins bound to 20 µg Dynabeads were eluted in 2 µl 0.1% TFA, 5 mM N-octylglycoside and, after addition of 1 µl 100 mM $(NH_4)HCO_3$, pH 7.8, digested on an immobilized-trypsin microcolumn followed by MALDI-TOF MS analysis as described[4]. Database searches were performed with the software developed by MS-Fit provided by K. Clauser and P. Baker (Univ. of California, San Francisco Mass Spectromerty Facility, San Francisco, CA; available at http://prospector.ucsf.edu). For additional data processing, the software package GPMAW.3.12, Lighthouse data, Odense, Denmark, was used. Bands excised from SDS–polyacrylamide gels were subjected to mass spectrometric peptide mapping as described[22,23].

Acknowledgments

This work was supported by grants from the Commission of the European Communities (TMR, ERBFMBICT950446) and the Danish Biotechnology Program.

1. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246 (1989).
2. Rebar, E.J. & Pabo, C.O. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671–673 (1994).
3. Karas, M. & Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 Daltons. *Anal. Chem.* **60**, 2299–2301.
4. Gobom, J., Nordhoff, E., Ekman, R. & Roepstorff, P. Rapid micro-scale proteolysis of proteins for MALDI-MS peptide mapping using immobilized trypsin. *Int. J. Mass Spectrom. Ion Proc.* **169/170**, 153–163 (1997).
5. Henzel, W.J., Billeci, T.M., Stults, J.T. & Wong, S.C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA* **90**, 5011–5015 (1993).
6. Parkinson, G. et al. Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. *J. Mol. Biol.* **260**, 395–408 (1996).
7. Kolb, A., Busby, S., Buc, H., Garges, S. & Adhya, S. Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* **62**, 749–795 (1993).
8. Valentin-Hansen, P., Svenningsen, B.A., Munch-Pedersen, A. & Hammer-

Jespersen, K. Regulation of the *deo* operon in *Escherichia coli*: the double negative control of the *deo* operon by the cytR and deoR repressors in a DNA directed in vitro system. *Mol. Gen. Genet.* **159**, 191–202 (1978).
9. Heyduk, T. & Lee, J.C. *Escherichia coli* cAMP receptor protein: evidence for three protein conformational states with different promoter binding affinities. *Biochemistry* **28**, 6914–6924 (1989).
10. Allegretto, E.A. et al. Transactivation properties of retinoic acid and retinoid X receptors in mammalian cells and yeast. Correlation with hormone binding and effects of metabolism. *J. Biol. Chem.* **268**, 26625–26633 (1993).
11. Henry, K., O'Brien, M.L., Clevenger, W., Jow, L. & Noonan, D.J. Peroxisome proliferator-activated receptor response specificities as defined in yeast and mammalian cell transcription assays. *Toxicol. Appl. Pharmacol.* **132**, 317–324 (1995).
12. Lechler, A. & Kreutzer, R.. The phenylalanyl-tRNA synthetase specifically binds DNA. *J. Mol. Biol.* **278**, 897–901 (1998).
13. Um, S. et al. Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem.* **273**, 20728–20736 (1998).
14. Tontonoz, P., Spiegelman, B.M. & Hu, E. Stimulation of adipogenesis in fibroblasts by PPAR γ 2, a lipid-activated transcription factor. *Cell* **79**, 1147–1156 (1994).
15. Wakabayashi-Ito, N. & Nagata S. Characterization of the regulatory elements in the promoter of the human elongation factor-1 α gene. *J. Biol. Chem.* **269**, 29831–29837 (1994).
16. Satoh, M. & Lindahl, T. Role of poly(ADP-ribose) formation in DNA repair. *Nature* **356**, 356–258 (1992).
17. Tewari, M. et al. Yama/CPP32 β, a mammalian homolog of CED-3, is a CrmA-inhibitable protease that cleaves the death substrate poly(ADP-ribose) polymerase. *Cell* **81**, 801–809 (1995).
18. Meisterernst, M., Stelzer, G. & Roeder, R.G.. Poly(ADP-ribose) polymerase enhances activator-dependent transcription in vitro. *Proc. Natl. Acad. Sci. USA* **94**, 2261–2265 (1997).
19. Zhang, X., Gunasekera, A., Ebright, Y.W. & Ebright, R.H. Derivatives of CAP having no solvent-accessible cysteine residues, or having a unique solvent-accessible cysteine residue at amino acid 2 of the helix-turn-helix motif. *J. Biomol. Struct. Dyn.* **9**, 463–473 (1991).
20. Dent, C.L. & Latchman, D.S. Transcription factors, a practical approach (Oxford University Press, Oxford; 1993).
21. Dignam, J.D., Lebovitz, R.M. & Roeder, R.G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic. Acids. Res.* **11**, 1475–1489 (1983).
22. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
23. Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R. & Roepstorff, P. Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass. Spectrom.* **34**, 105–116 (1999).