Combination of MS Protein Identification and Bioassay of Chromatographic Fractions to Identify Biologically Active Substances from Complex Protein Sources

Sadao Kuromitsu‡, Hiroyuki Yokota‡§, Masashi Hiramoto‡, Masatoshi Yuri‡, Masanori Naitou‡, Naoto Nakamura‡, Shigeki Kawabata‡, Masato Kobori‡, Masao Katoh‡, Kiyoshi Furuchi‡, Haruhisa Mita¶, and Tetsuo Yamada¶

Purification of biologically active proteins from complex biological sources is a difficult task, usually requiring large amounts of sample and many separation steps. We found an active substance in a serum response element-dependent luciferase reporter gene bioassay in interstitial cystitis urine that we attempted to purify with column chromatography and the bioassay. With anion-exchange Mono Q and C₄ reversed-phase columns, apparently sharp active peaks were obtained. However, more than 20 kinds of proteins were identified from the active fractions with MS, indicating that the purification was not complete. As further purification was difficult, we chose a candidate molecule by means of studying the correlation between MS protein identification scores and bioassay responses of chromatographic fractions near the active peaks. As a result, epidermal growth factor (EGF) was nominated as a candidate molecule among the identified proteins because the elution profile of EGF was consistent with that of the bioassay, and the correlation coefficient of EGF between MS protein identification scores and bioassay responses was the highest among all the identified proteins. With recombinant EGF and anti-EGF and anti-EGF receptor antibodies, EGF was confirmed to be the desired substance in interstitial cystitis urine. This approach required only 20 ml of urine sample and two column chromatographic steps. The combination of MS protein identification and bioassay of chromatographic fractions may be useful for identifying biologically active substances from complex protein sources. Molecular & Cellular Proteomics 8:1318–1323, 2009.

Nevertheless the recent progress of MS has dramatically changed protein analysis (4). With MS, smaller protein samples can be used than with classical protein identification methods such as N-terminal peptide sequencing.

Interstitial cystitis (IC)¹ is a chronic inflammatory disease characterized by frequency and urgency and/or severe pelvic pain (5). The International Continence Society also selected the term “painful bladder syndrome” for IC (6). The quality of life of IC patients is extremely low because of their severe symptoms. The pathogenesis of IC is unclear, and effective treatments have not been established. To elucidate the mechanism of IC pathogenesis, we attempted to find characteristic proteins in IC urine using proteomics techniques and have already reported active neutrophil elastase as an IC urinary marker (7). We had also performed gene expression analysis of IC bladder tissues using GeneChip technology and found that mRNA expression of GPR18, a member of the G-protein-coupled receptors, was higher in IC bladder than in the control.² We tried to confirm whether GPR18 endogenous ligand existed in IC urine by using a bioassay with GPR18 transfected cells.

In the present study, the existence of an active substance in IC urine was suggested in the bioassay using the serum response element (SRE)-dependent luciferase reporter gene with the stable recombinant HEK293 cell line expressing GPR18. We thought that the response was derived from GPR18 and tried to purify the active substance from a small volume of IC urine using chromatographic techniques. Among the many proteins identified from partially purified samples, we clearly nominated epidermal growth factor (EGF) as a candidate molecule judging from the correlation between MS protein identification and the bioassay of chromatographic fractions. With recombinant EGF and anti-EGF antibody, EGF was confirmed to be the desired substance found in IC urine.

¹ The abbreviations used are: IC, interstitial cystitis; SRE, serum response element; EGF, epidermal growth factor.

² S. Kuromitsu, H. Yokota, M. Hiramoto, M. Yoshino, K. Tsunoyama, S. Kawabata, M. Kobori, M. Katoh, K. Furuchi, H. Mita, and T. Yamada, manuscript in preparation.
The complete inhibition of the bioassay response by anti-EGF receptor antibody also indicated that the response was based on the EGF receptor, not GPR18, suggesting that GPR18 overexpression enhanced the EGF signal via the endogenous EGF receptor of the HEK293 cell line.

EXPERIMENTAL PROCEDURES

Materials and Reagents—Sequencing grade modified trypsin was purchased from Promega Co. (Madison, WI), Vydac C4 (0.46-cm inner diameter x 15 cm) was purchased from the Separations Group (Hesperia, CA), Sep-Pak C18 and Rapigest SF were purchased from Waters (Milford, MA), Mono Q HR 5/5 (0.5-cm inner diameter x 5 cm) was purchased from GE Healthcare, recombinant human EGF was purchased from PeproTech Inc. (Rocky Hill, NJ), anti-human EGF antibody was purchased from R&D Systems, Inc. (Minneapolis, MN), anti-EGF receptor antibody was purchased from EMD Biosciences, Inc. (La Jolla, CA), PepMap C18 cartridge (0.3-mm inner diameter x 5 mm; 5 μm) was purchased from LC Packings (Amsterdam, Netherlands), nano-HPLC capillary column (0.075-mm inner diameter x 150 mm; C18; 5 μm) was purchased from Nikkyo Technos (Tokyo, Japan), pSRE (serum response element)-luciferase reporter plasmid was purchased from Stratagene (La Jolla, CA), and Lipofectamine 2000 and Dulbecco’s modified Eagle’s medium were purchased from Invitrogen. Human spleen cDNA and pSRE-luciferase reporter plasmid were purchased from Clontech. pEF-BOS-neo vector (8) was donated by Prof. S. Nagata (Osaka University Medical School, Osaka, Japan). All other reagents were of analytical grade.

IC Patient—The 31 IC patients satisfied the National Institute of Diabetes and Digestive and Kidney Diseases criteria (9). The mean age of the IC patients was 58 years (range, 19–82 years); 84% were female. The mean age of the 11 healthy volunteers without complaints of lower urinary symptoms was 42 years (range, 20–60 years); 82% were female. The patients and healthy volunteers were enrolled in accordance with the research plan approved by the ethics committee of both the National Hospital Organization Sagamihara National Hospital and Astellas Pharma Inc. and gave written informed consent.

Purification of an Active Substance from IC Urine—Random midstream urine samples were collected in 15- or 50-ml polypropylene conical tubes and immediately centrifuged at 3,000 rpm for 10 min at 4 °C. The supernatant was immediately stored in the same kind of tube without any additives at −20 °C. The maximum storage period and the maximum number of freeze-thaw cycles of the urine samples were 6 years and 10 cycles, respectively. Just before use, the urine was thawed and immediately centrifuged at 3,000 rpm for 10 min at 4 °C. For purification, 20 ml of the supernatant was applied to a reversed-phase Sep-Pak C18 cartridge prewashed sequentially with 90:10 (v/v) acetonitrile/water and with water. The cartridge was
washed with water and then eluted with 50:50 (v/v) acetonitrile/water. This fraction was concentrated with a SpeedVac concentrator (Thermo Electron, Waltham, MA), cold ethanol was added, and the mixture was stored at 4 °C for more than 30 min. After centrifugation at 21,000 g for 5 min, the precipitate was dried with the SpeedVac concentrator, dissolved with 0.25 ml of 20 mM Tris-HCl (pH 8.0), and then injected on an anion-exchange column (Mono Q HR 5/5). The mobile phases consisted of Solvent A (20 mM Tris-HCl, pH 8.0) and Solvent B (20 mM Tris-HCl, pH 8.0, 1 M NaCl). The column was initially equilibrated at 0% B at a flow rate of 0.5 ml/min. The separation was performed by a linear gradient of 0% B to 100% B over 30 min at a flow rate of 1 ml/min at room temperature. Detection was by UV absorbance at 280 nm. The fractions were manually collected at 1-min intervals into polypropylene microtubes, which were used for the bioassay and MS protein identification.

The Mono Q fractions were further separated with a reversed-phase column (Vydac C4). The mobile phases consisted of Solvent A (0.1% TFA) and Solvent B (0.1% TFA in 90:10 (v/v) acetonitrile/water). The column was initially equilibrated at 0% B at a flow rate of 0.3 ml/min. The separation was performed by a linear gradient of 0% B to 50% B over 30 min at a flow rate of 1 ml/min at room temperature. Detection was by UV absorbance at 280 nm. The fractions were manually collected at 1-min intervals into polypropylene microtubes, which were used for the bioassay and MS protein identification. The HPLC experiments were performed using an HP1090M liquid chromatograph (Agilent Technologies, Palo Alto, CA).

**MS Analysis**—The collected fraction was dried with the SpeedVac concentrator and dissolved with 50 µl of 50 mM NH₄HCO₃ and 0.1% Rapigest SF. The reduction and alkylation were performed with 5 mM DTT at 60 °C for 30 min and with 15 mM iodoacetamide at ambient temperature for 20 min. After 1 µg of trypsin was added, the reaction mixture was incubated at 37 °C for 1 h, then HCl was added at a final concentration of 0.1 M, and the mixture was centrifuged at 13,000 rpm for 10 min to remove Rapigest SF. The supernatant, digested peptide mixture was concentrated on an on-line trap column (PepMap C₁₈) and then separated with a reversed-phase microcapillary column (nano-HPLC capillary column) using a CapLC system (Waters). The mobile phases consisted of Solvent A (2% acetonitrile, 0.1% formic acid) and Solvent B (98% acetonitrile, 0.1% formic acid).
acid) and Solvent B (90% acetonitrile, 0.1% formic acid). The column was initially equilibrated at 10% B at a flow rate of 200 nl/min. The separation was performed by a linear gradient of 10% B to 50% B over 35 min, 90% B over 10 min, and 10% B over 15 min at a flow rate of 200 nl/min at room temperature. The LC effluent was directly interfaced with the nanoelectrospray ion source on a Q-TOF Ultima API mass spectrometer (Waters). An MS survey scan was obtained for the m/z range of 400–1500, and MS/MS spectra were acquired for the two most intense ions from the survey scan with precursor charge limitation (2 or greater). Dynamic exclusion of 2-min duration was used to acquire MS/MS spectra from low intensity ions. All the MS/MS spectra were converted into text files (.pkl) using Protein Lynx software (Waters; purchased in February 2003) combining sequential scans with the same precursor, smoothing the spectrum with Savitzky-Golay smoothing, and measuring the peak top with a centroid top of 80%.

Protein identification was performed using MASCOT software (Version 2.1.0, Matrix Science Inc., Boston, MA). An in-house customized database (built in September 2006; 224,925 sequences) based on the NCBI non-redundant protein sequence database with species limitation (only human, mouse, rat, and bovine proteins can pass) and with locus redundancy removal by NCBI EntrezGene was used. The MASCOT search parameters were as follows. Peptide tolerance was 0.45 Da, and MS/MS tolerance was 0.15 Da (monoisotopic mass). Fixed modification of carbamidomethylation (Cys) and variable modifications of oxidation (Met), acetylation (N terminus), and pyro-Glu (Glu and Gln) were selected, and up to four missed trypsin cleavages were allowed. A protein score of 40 (p < 0.05) and a peptide score of 25 were the minimum identification criteria, and manual examination was conducted for all proteins identified with fewer than 80 points on the protein score or fewer than four unique peptides. The criteria used for manual validation included (a) peptide fragment ions being clearly above base-line noise, (b) intense y or b ions corresponding to the N-terminal cleavage of proline or glycine and absence of C-terminal cleavage of proline or glycine (except where proline follows), (c) major five fragment ions being interpretable, and (d) not matching a common contaminant (such as silicone or polyethylene glycol) or tryptic-derived signals. In addition, each MS/MS spectrum of proteins identified with a small number (one or two) of peptides was compared with the same spectrum of the same proteins (if it exists) identified using a large set (three or more) of peptides for further manual confirmation. If peptides matched to multiple database entries, the one RefSeq accession number belonging to the species Homo sapiens was assigned for the identification protein.

Bioassay—The full-length cDNA of GPR18 (GenBank™ accession number NM_005292) was amplified by PCR with forward primer 5’-GGGTCTAGAATGATCACCCTGAACAATC-3’ and reverse primer 5’-GGGTCTAGATCATAACATTTCACTGTTTATA-3’ from human spleen cDNA and ligated into pEF-BOS-neo vector treated with Xbal. The stable recombinant HEK293 cell lines expressing GPR18 and vector control were established by selecting G418-resistant cells after transfection with the pEF-BOS-neo-GPR18 vector and the pEF-BOS-neo vector, respectively. The cells were seeded onto collagen-coated 96-well plates (10,000 cells/well) in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum, 100 units/ml penicillin, and 100 μg/ml streptomycin and incubated at 37 °C in a 5% CO2 humidified atmosphere for 24 h and then transfected with pSRE-luciferase reporter plasmid (100 ng/well) using Lipofectamine 2000. After 24 h, 10 μl of urine samples were added into the wells and incubated for 16 h. After discarding the medium, the cells were lysed in 20 μl of solubilization buffer (0.5 mM MgCl2, 10 mM dithiothreitol, 0.1% Triton X-100, 10 mM Tris, pH 7.8), and then 100 μl of substrate solution (5 mM luciferin, 2 mM coenzyme A, 2 mM ATP, 0.5 mM MgCl2, 2 mM Mg(OH)2, 10 mM Tris, pH 7.8) were added into the wells. Luciferase activity measured by a Dynatech Mls3000 luminometer (Dynatech Laboratories, Chantilly, VA) is expressed as relative light units.
RESULTS

We found the existence of an active substance in IC urine in the bioassay using the SRE-dependent luciferase reporter gene with the stable recombinant HEK293 cell line expressing GPR18. The results are shown in Fig. 1. IC urine showed higher activity than healthy volunteer urine (Fig. 1A). The signal with the HEK293 cell line with GPR18 gene-expressing vector was higher than one with vector control (Fig. 1B), suggesting that the activity was related to the GPR18 gene. The activity was lost with protease digestion or heat treatment (data not shown), indicating that this substance is a polypeptide.

We tried to purify it using column chromatography. The scheme of purification is shown in Fig. 2. With Mono Q anion-exchange chromatography, two main active peaks were obtained (Fig. 3A). We collected the two active fractions designated as AEX-12 and AEX-14, which were further separated, respectively, with C4 reversed-phase chromatography. From the AEX-12 fraction, one active peak was obtained (Fig. 3B). The fraction was dried, digested with trypsin, and then identified with MS. Although the peak seemed very sharp, more than 10 kinds of proteins were detected in the active fraction, number 34, from AEX-12 (Table I). Whereas from the AEX-14 fraction, one relatively broad active peak eluted at three fractions was obtained (Fig. 3C). More than 20 kinds of proteins were also identified in the three active fractions, numbers 34–36, from AEX-14 (Table I). We attempted to purify the active fractions using additional size exclusion, cation-exchange, or lectin affinity chromatography but failed to achieve satisfactory purification.

As further purification was difficult, we tried another strategy. The fractions eluted near the active peaks on reversed-phase chromatography were identified with MS, and then the elution profile of the proteins was compared with that of the bioassay responses. The results are shown in Table I. EGF was identified only in the active fractions of both number 34 of AEX-12 and numbers 34–36 of AEX-14 (Table I). The production spectrum of the identified EGF was identical with that of authentic recombinant EGF (Fig. 4). The correlation coefficient of EGF between MASCOT scores and bioassay responses was the highest among those of the identified proteins (Table I). Therefore, we nominated EGF as the most prominent candidate molecule.

Fig. 5. The activity of IC urine was inhibited with anti-EGF and anti-EGF receptor antibodies. After transfection of pSRE-luciferase plasmid into GPR18-expressing HEK293 cells, EGF (A) and IC urine (B) were added into wells, and luciferase activity was measured. After anti-EGF antibody was added into wells, 0.2 ng of EGF (C) and 10 μl of IC urine (D) were assayed. After anti-EGF receptor (EGFR) antibody was added into wells, 0.2 ng of EGF (E) and 10 μl of IC urine (F) were assayed. The bar indicates the S.E. (n = 3). RLU, relative light units; hEGF, human EGF.
To confirm whether the desired active substance in IC urine was EGF, we studied the effect of authentic recombinant EGF protein and anti-EGF and anti-EGF receptor antibodies for the bioassay. The recombinant EGF showed a dose-dependent response in the bioassay as well as IC urine (Fig. 5, A and B). The concentration of EGF in the IC urine sample used for this purification was estimated to be ~10 ng/ml because the signals of 1 and 10 μl of IC urine corresponded to those of 0.01 and 0.1 ng of EGF, respectively. The activities of recombinant EGF and IC urine were completely inhibited with 0.01 and 0.1 ng of EGF, respectively. The activities of recombinant EGF and IC urine were completely inhibited with anti-EGF and anti-EGF receptor antibodies (Fig. 5, C–F). These results show that the active substance existing in IC urine was EGF.

**DISCUSSION**

EGF is a 53-amino acid peptide with three internal disulfide bonds and has a variety of biological activities (2, 3). EGF in urine is reported to have C-terminal heterogeneity such as the desarginylation form (2, 3), which should be the cause of the multiple peaks observed with Mono Q chromatography. It has also been reported that the average EGF concentration in IC urine is higher than in the control at 20 ng/ml versus 2 ng/ml, respectively (10); this was consistent with our data that the activity derived from EGF in IC urine was higher than that in the control. In the previous reports, EGF was purified from a large volume of human urine with many chromatographic steps such as 330 liters with four steps (2) and 500 liters with six steps (3). In this report, EGF was identified as the active substance from 20 ml of IC urine with only two chromatographic steps.

A decrease in necessary sample amounts for protein identification was achieved with MS, but at the same time, many kinds of proteins were identified from a partially purified sample, making it difficult to choose a candidate molecule to be validated among them. This report shows that it is possible to select a suitable candidate molecule among many proteins identified with MS by judging from the correlation between MS protein identification and bioassay of chromatographic fractions. This approach does not need complete purification of the molecule. Although this methodology was not proven to be available for other samples such as serum in this report and should not be directly available in the case where plural components are necessary for bioassay activation, MS protein identification for monitoring chromatographic fractions as well as bioassay may be useful for other similar cases.

The SRE signal of EGF and IC urine was also obtained with the HEK293 cell line with vector control to a lesser extent than that with the GPR18 gene, which was completely inhibited by the anti-EGF receptor antibody (data not shown). Before identification of EGF, we believed that the response with control transfectant was due to endogenous GPR18 of HEK293 cells. However, with recombinant EGF and anti-EGF antibody, EGF was confirmed to be the desired substance found in IC urine, and the complete inhibition of the bioassay response by anti-EGF receptor antibody indicated that the response was based on the EGF receptor, not GPR18. The reason why the response with the GPR18-expressing HEK293 cell line was higher than that with the control is unclear. The SRE signal of EGF via the endogenous EGF receptor of the HEK293 cell line might be enhanced by overexpressing the GPR18 gene; this remains to be elucidated.

In this report, the elution profile of proteins with MS identification was evaluated with semiquantitative MASCOT identification scores. The use of quantitative MS protein analysis with bioassay in column chromatography will make it much easier to nominate candidate proteins from low amounts of crude protein samples with minimum purification steps.

**Acknowledgments**—We thank Mari Masumoto, Akiko Ohtsu, Misaki Mito, and Kumiko Fukumaru for technical assistance.

[5] The on-line version of this article (available at http://www.mcpjournal.org) contains supplemental material.

[§ To whom correspondence should be addressed. Tel.: 81-29-852-5111; Fax: 81-29-852-5444; E-mail: hiroyuki.yokota@jp.astellas.com.

**REFERENCES**

1. Cutler, P. (2004) *Protein Purification Protocols*, 2nd Ed., Humana Press Inc., Totowa, NJ

2. Mount, C. D., Lukas, T. J., and Orth, D. N. (1985) Purification and characterization of epidermal growth factor (β-urogastrone) and epidermal growth factor fragments from large volumes of human urine. *Arch. Biochem. Biophys.* **240**, 33–42

3. Jaspar, J. M., and Franchimont, P. (1987) Purification of human epidermal growth factor (urogastrone) from urine. *Eur. J. Biochem.* **166**, 295–301

4. Domon, B., and Aebersold, R. (2006) Mass spectrometry and protein analysis. *Science* **312**, 212–217

5. Ebadawi, A. (1997) Interstitial cystitis: a critique of current concepts with a new proposal for pathologic diagnosis and pathogenesis. *Urology* **49**, 14–40

6. Abrams, P., Cardozo, L., Fall, M., Griffiths, D., Rosier, P., Ulmsten, U., van Kesterenbroek, P., Victor, A., and Wein, A. (2002) The standardisation of terminology of lower urinary tract function: report from the standardisation sub-committee of the International Continence Society. *Neuromuscul. Urodyn.* **21**, 167–178

7. Kuromitsu, S., Yokota, H., Hiramoto, M., Morita, S., Mita, H., and Yamada, T. (2008) Increased concentration of neutrophil elastase in urine from patients with interstitial cystitis. *Scand. J. Urol. Nephrol.* **42**, 455–461

8. Mizushima, S., and Nagata, S. (1990) PEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res.* **18**, 5322

9. Gillenwater, J. Y., and Wein, A. J. (1987) Summary of the National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases Workshop on Interstitial Cystitis, National Institutes of Health, Bethesda, Maryland, August 28–29, J. Urol. **140**, 203–206

10. Zhang, C. O., Li, Z. L., Shoenfelt, J. L., Kong, C. Z., Chai, T. C., Erickson, D. R., Peters, K. M., Rovner, E. S., and Keay, S. (2003) Comparison of APF activity and epithelial growth factor levels in urine from Chinese, African-American, and white American patients with interstitial cystitis. *Urology* **61**, 897–901

11. Zieske, L. R. (2006) A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J. Exp. Bot.* **57**, 1501–1508