# Proteomic identification of collagens and related proteins in human fibroblasts

## **J. E. Oh<sup>1</sup>, K. Krapfenbauer<sup>2</sup>, and G. Lubec<sup>1</sup>**

<sup>1</sup> Department of Pediatrics, University of Vienna, Vienna, Austria <sup>2</sup> Center for Medical Genomics, F. Hoffmann-La Roche, Basel, Switzerland

Received March 4, 2004 Accepted May 3, 2004 Published online October 4, 2004;  $\odot$  Springer-Verlag 2004

Summary. Fibroblasts are used for diagnosis of a series of metabolic diseases and are particularly suitable for the diagnosis of collagen disorders. We aimed to generate a skin fibroblast map that would be suitable for the concomitant determination of collagen and collagen-related proteins.

A human skin fibroblast cell line was cultivated, homogenised, proteins extracted and subject to two-dimensional gel electrophoresis with subsequent in-gel-digestion of protein spots and mass spectrometrical identification (MALDI-TOF).

Collagen alpha1 (I) chain precursor, collagen alpha1 (III) chain precursor, collagen alpha2 (VI) precursor and collagen modifying enzymes prolyl 4-hydroxylase alpha-2-subunit precursor, procollagen-lysine 2-oxoglutarate 5-dioxygenase 1 and 2, protein disulfide isomerase ER-60 precursor and peptidyl-prolyl cis-trans isomerase were among the abundant proteins.

The finding of collagen and collagen-related structures as well as the identification of other metabolic enzyme systems on one 2D gel may propose the use of this proteomic method for further characterization of collagen and collagen-related proteins or for preliminary screening of metabolic disorders.

Keywords: Fibroblast – Collagen – Collagen alpha1 chain – Procollagenlysine – Prolyl 4-hydroxylase – Protein disulfide isomerase

#### Introduction

Collagens, collagen-related proteins and collagen disorders are still a major focus of research on hereditary and acquired connective tissue diseases including Ehlers Danlos syndrome, osteogenesis imperfecta, chondrodysplasias, Alport syndrome, osteoarthrosis, arterial aneurysms, to name a few (Myllyharju and Kivirikko, 2001). The underlying mechanisms for inherited collagen deficits are mainly mutations of collagen chains (Kuivaniemi et al., 1997) or of enzymes mediating genetically determined posttranslational modifications (Yeowell and Walker, 2000).

Molecular diagnosis confirming clinical diagnosis is a domain of molecular biology techniques (Kuivaniemi et al., 1997) and although electrophoresis is used to screen collagen chains or involves specific antibodies for immunoblotting, no protein chemical method for the use of collagens in skin fibroblasts, independent of antibody specificity and availability, has been reported so far.

Leung and coworkers (2001) generated a reference map of human lung fibroblast proteins consisting of 98 protein spots including procollagen-proline and a prolyl 4-hydroxylase subunit using immobilized pH gradient–isoelectric focusing-based two-dimensional gel electrophoresis with subsequent mass spectrometrical analysis.

In the present study we aimed to concomitantly identify and characterize several collagen chains and two major collagen related hydroxylases, lysylhydroxylase and prolylhydroxylase (Royce et al., 1989) using a comparable proteomic technique. In addition, we decided to identify abundant fibroblast proteins including enzymes from different metabolic pathways and cascades in order to provide an analytical tool to screen for corresponding metabolic deficiencies.

This method would be suitable for preliminary screening collagen and metabolic disorders and could be extended to MS–MS analysis for sequencing the gene products in addition to genes as it is not known whether collagen mutations do result into deficient collagens at the protein level.

The concomitant determination of reference proteins housekeeping proteins as e.g. actin, tubulin and glyceraldehyde 3 phosphate dehydrogenase allows normalization versus housekeeping proteins on the same gel. Last not least we attempted to find different expression forms, probably reflecting isoforms or posttranslational modifications, of these structures.

#### Materials and methods

#### Fibroblast cell cultivation

A human skin fibroblast cell line was obtained from Prof. M. Hengstschläger, Dpt of Obstetrics and Gynecology, Prenatal Diagnosis and Therapy, University of Vienna and cultured in Dulbecco's modified Eagle's medium (Life Technologies, Inc.) supplemented with 10% FBS (PAA Laboratories, Linz, Austria), 4 mM glutamine (Biochrom, Berlin, Germany) and 25 mM glucose.

#### Sample preparation

Harvested cells were washed three times in 10 mL phosphate buffered saline (Gibco BRL), centrifuged for 10 min at 800 g at room temperature and subsequently homogenised with 1.0 ml of sample buffer consisting of 7 M urea (Merck, Germany), 2 M thiourea (Sigma, st. Louis, MO), 4% CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propane-sulfonate) (Sigma), 65 mM 1,4-dithioerythritol (Merck), 1 mM EDTA (ethylenediaminetraacetic acid) (Merck), 1 mM PMSF, 0.5% carrier ampholytes and protease inhibitor complete (Roche, Switzerland). After homogenization samples were left at room temperature for 1 h and centrifuged at 14,000 rpm for 60 min and the supernatant was transferred into Ultrafree-4 centrifugal filter unit (Millipore, Bedford, MA), for desalting and concentrating proteins. Protein content of the supernatant was quantified by Bradford protein assay system (Bradford, 1976). The standard curve was generated using bovine serum albumin and absorbance was measured at 595 nm.

#### Two dimensional gel electrophoresis (2-DE)

Samples prepared from each cell line were subjected to 2-DE as described elsewhere (Langen et al., 1999; Weitzdoerfer et al., 2002). 1 mg protein was applied on immobilised pH 3–10 nonlinear gradient strips in sample cups at their basic and acidic ends. Focusing was started at 200 Vand the voltage was gradually increased to 8000 V at  $4 \text{ V/min}$  and kept constant for a further 3 h (approximately 150000 Vh totally). After the first dimension, strips (13 cm) were equilibrated for 15 min in the buffer containing 6 M urea, 20% glycerol, 2% SDS, 2% DTT and then for 15 min in the same buffer containing 2.5% iodoacetamide instead of DDT. After equilibration, strips were loaded on 9–16% gradient sodium dodecylsulfate polyacrylamide gels for seconddimensional separation. Gels  $(180 \times 200 \times 1.5 \text{ mm})$  were run at 40 mA per gel. Immediately after the second dimension run, gels were fixed for 12 h in 50% methanol containing 10% acetic acid and stained with colloidal Coomassie blue (Novex, San Diego, CA) for 12 h on a rocking shaker. Molecular masses were determined by running standard protein markers (Bio-Rad Laboratories, Hercules, CA), covering the range 10–250 kDa. pI values were used as given by the supplier of the immobilized pH gradient strips (Amersham Bioscience, Uppsala, Sweden). Excess of dye was washed out from the gels with distilled water and gels were scanned with ImageScanner (Amersham Bioscience, Uppsala, Sweden). Electronic images of the gels were recorded using Photoshop (Adobe) and PowerPoint (Microsoft) software.

#### Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS)

MALDI-MS analysis was performed as described (Berndt et al., 1999; Krapfenbauer et al., 2001) with some modifications. Spots were excised with a spot picker and gel pieces were washed twice in  $100 \mu l$  of  $30\%$ acetonitrile in 50 mM ammonium bicarbonate and dried in a Speedvac evaporator for 10 min. Proteins were rehydrated with  $4 \mu$ l of 3 mM Tris-HCl, pH 9.0, containing 50 ng trypsin (Promega, Madison, WI) for 16 h or

overnight at 37°C. Peptide extracts were vacuum-dried and resuspended in  $7 \mu$ l of distilled water and shaken for 10 min.  $4 \mu$ l of 50% acetonitrile; containing 0.3% trifluoroacetic acid and the standard peptides, des-Argbradykinin (Sigma, 2465.1989 Da); were added to each gel piece and shaken for 10 min. Sample application was performed using SymBiot I sample processor (PE Biosystems, Framingham, MA).  $1.5 \mu$ l of digested peptide extracts were simultaneously spotted onto a MALDI target in  $1 \mu$  of matrix, consisting of a saturated solution of alpha-cyano-4-hydroxycinnamic acid (Sigma) in 50% acetonitrile containing 0.1% trifluoroacetic acid. MALDI-MS was performed using a Reflex III reflector time-of-flight mass spectrometer (Reflex 3, Bruker Analytics, Bremen, Germany). An accelerating voltage of 20 kV was used. Peptide matching and protein searches were performed automatically. The peptide masses were compared with the theoretical peptide masses of all available proteins from all species. Monoisotopic masses were used and a mass tolerance of 0.0025% was allowed. The algorithm used for determining the probability of a false positive match with a given MS-spectrum is described elsewhere (Berndt et al., 1999).

#### Results

A series of 80 proteins (SWISSPROT numbers indicated) has been unambiguously identified by MALDI-TOF and grouped into the categories: collagen related proteins, structural proteins, cytoskeleton related proteins, metabolic proteins, transcription and translation related proteins, proteasome related proteins, antioxidant proteins and miscellaneous proteins.

Figure 1 presents the map of identified and assigned proteins in fibroblast cell.



Fig. 1. Two-dimensional map of a whole cell lysate extracted from cultivated fibroblast cell line. The proteins were extracted and separated on pH 3–10 nonlinear IPG strip, followed by 9–14% SDS-polyacrylamide gel. The gel was stained with Coomassie blue. The spots were analyzed by MALDI-MS and the names of identified proteins are listed in Table 1

## Table 1. The list of identified proteins in fibroblast cell lines



(continued)

### Table 1 (continued)



(continued)

## Table 1 (continued)





Enzyme	EC number	Event catalyzed	Substrate requirement
Signal Peptidase		Cleavage of signal peptide of pre-pro $\alpha$ chains	Wide range of unrelated newly synthesized proteins
Prolyl 4-hydroxylase	1.14.11.2	4-Hydroxylation of proline	-X-Pro-Gly-
Prolyl 3-hydroxylase	1.14.11.7	3-Hydroxylation of proline	-Pro-4Hyp-Gly-
Lysyl hydroxylase	1.14.11.4	Hydroxylation of lysine	-X-lys-Gly-
Hydroxylysyl galactosyl-transferase	2.4.1.50	O-Glycosylation of hydroxylysine	UDP-galactose Hydroxylysine in peptide linkage
Hydroxylysyl glucosyl-transferase	2.4.1.66	O-Glycosylation of galactosyl hydroxylysine	UDP-glucose Galactosyl-hydroxylysine in peptide linkage
Protein disulphide isomerase	5.3.4.1	Native disulphide bond formation by thiol: disulphide interchange	Newly synthesized polypeptide chains
Prolyl-peptidyl cis/trans isomerase		Interconversion of <i>cis</i> and <i>trans</i> conformers of the peptide backbone around the planar imide bond in the sequence -X-Pro-	Newly synthesized polypeptide chains

Table 2. Collagen posttranslation related enzymes

Furthermore, fibroblast proteins were identified based upon molecular weight and pI value using 2-D gel electrophoresis and MALDI-MS analysis (Table 1).

While most proteins were represented by a single spot, some proteins including collagen and collagen-related proteins show the presence of more than one spot (Fig. 1). For example, collagen alpha 1(I) chain precursor shows two spots with same molecular weight but different pI values (Fig. 1, Table 1). The presence of several spots assigned to the identical protein may indicate posttranslational modifications or isoforms. Accession numbers and protein names are listed in Table 1.

Table 2 shows the collagen post-translational related enzymes with the related events and the required substrates (Kielty et al., 1993). Prolyl 4-hydroxylase, Protein disulfide isomerase and Peptidyl-prolyl cis-trans isomerase are relevant to the proteins observed in our data (Table 1).

#### **Discussion**

The results shown herein allow identification, characterisation of three collagen chains that are deficient in collagen disorders as well as two key enzymes responsible for genetically determined hydroxylation of collagenous proline and lysine representing posttranslational modifications. Collagen hydroxylation deficits are major causes of inherited as well as acquired collagen disorders (Wenstrup et al., 1989). A major finding of the present study is also the possibility to determine two out of three known forms (Rautavuoma et al., 2002) of procollagen-lysine, 2-oxoglutarate 5-dioxygenases 1 and 2 on the same gel. Leung and coworkers (2001) already reported a lung fibroblast map showing the presence of collagen structures including Procollagen-proline, Collagen-binding protein 2 precursor, Prolyl 4-hydroxylase alpha-1 subunit precursor, Peptidyl-prolyl cis-trans isomerase A, but did not observe collagen chains.

Moreover, a series of metabolic enzymes provides the possibility of studying metabolic derangement at the protein level.

Many proteins including collagen and collagen-related proteins were presented by multiple spots and these may reflect isoforms or posttranslational modifications. Collagen alpha 1(I) chain precursor e.g. was showing two spots with comparable molecular weight but different pI values probably reflecting either lysine-hydroxylation and/or or O-Glycosylation of hydroxylysine (Table 1, Fig. 1).

In collagen disorders electrophoretic shifts can be expected and were already observed (Nuytinck et al., 2000): abnormal glycosylation, hydroxylation, truncation from procollagen to collagen can lead to different electrophoretic mobility and spots can then be even checked for mutations using MS–MS or sequencing, both techniques can be applied directly on the spot from 2DE (Fernandez et al., 2003).

The necessity of generating a map is given by the fact that proteins in a gel cannot be simply predicted by molecular weight and theoretical pI provided in databases or calculation methods (http://www.expasy.org). As shown in the table, in some cases observed pIs are divergent from the predicted ones and pIs of individual proteins may be different in individual cell lines and tissues and cannot be simply extrapolated.

The advantage of this proteomic method is the concomitant determination of a list of proteins permitting the generation of an expressional pattern and stoichiometrical evaluation (Fountoulakis, 2001). Moreover, the concomitant determination of several housekeeping proteins as e.g. cytoskeleton proteins and enzymes from carbohydrate metabolism can be used for normalisation of proteins versus these structures.

2-DE along with MALDI-MS of fibroblast proteins forms the analytical basis for subsequent studies to preliminary metabolic screens and further characterisation of isoforms and posttranslational modifications and mutations by additional advanced proteomic techniques. This proteomic approach was not designed to serve for medical diagnosis but rather for identification and characterisation of proteins forming the molecular basis of metabolic disorders.

#### Acknowledgment

We are highly indebted to the Red Bull Company, Salzburg, Austria, for generous financial support and we appreciate financial assistance of Verein zur Durchführung der wissenschaftlichen Forschung auf dem Gebiet der Neonatologie und Kinderintensivmedizin ''Unser Kind'', Vienna, Austria.

#### References

- Berndt P, Hobohm U, Langen H (1999) Reliable automatic identification from matrix-assisted laser desorption/ionization mass spectrometric peptide fingerprints. Electrophoresis 20: 3521–3526
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248–254
- Fernandez FM, Smith LL, Kuppannan K, Yang X, Wysocki VH (2003) Peptide sequencing using a patchwork approach and surface-induced dissociation in sector-TOF and dual quadrupole mass spectrometers. J Am Soc Mass Spectrom 14: 1387–1401
- Fountoulakis M (2001) Proteomics: current technologies and applications in neurological disorders and toxicology. Amino Acids 21: 363–381
- Kielty CM, Hophinson I, Grant ME (1993) Collagen: The collagen family: Structure, assembly, and organization in the extracellular

matrix. In: Royce PM, Steinmann B (eds) Connective tissue and its heritable disorder. Wiley Liss, Inc., pp 103–147

- Krapfenbauer K, Berger M, Friedlein A, Lubec G (2001) Changes in the levels of low-abundance brain proteins induced by kainic acid. Eur J Biochem 268: 3532–3537
- Kuivaniemi H, Tromp G, Prockop DJ (1997) Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. Hum Mutat 9: 300–315
- Langen H, Berndt P, Roder D, Cairns N, Lubec G, Fountoulakis M (1999) Two-dimensional map of human brain proteins. Electrophoresis 20: 907–916
- Leung KY, Wait R, Welson SY, Yan JX, Abraham DJ, Black CM, Pearson JD, Dunn MJ (2001) A reference map of human lung MRC-5 fibroblast proteins using immobilized pH gradient-isoelectric focusing-based two-dimensional electrophoresis. Proteomics 1: 787–794
- Myllyharju J, Kivirikko KI (2001) Collagens and collagen-related diseases. Ann Med 33: 7–21
- Nuytinck L, Freund M, Lagae L, Pierard GE, Hermanns-Le T, De Paepe A (2000) Classical Ehlers-Danlos syndrome caused by a mutation in type I collagen. Am J Hum Genet 66: 1398–1402
- Rautavuoma K, Takaluoma K, Passoja K, Pirskanen A, Kvist AP, Kivirikko KI, Myllyharju J (2002) Characterization of three fragments that constitute the monomers of the human lysyl hydroxylase isoenzymes 1–3. The 30-kDa N-terminal fragment is not required for lysyl hydroxylase activity. J Biol Chem 277: 23084–23091
- Royce PM, Moser U, Steinmann B (1989) Ehlers-Danlos syndrome type VI with normal lysyl hydroxylase activity cannot be explained by a defect in cellular uptake of ascorbic acid. Matrix 9: 147–149
- Weitzdoerfer R, Fountoulakis M, Lubec G (2002) Reduction of actinrelated protein complex  $2/3$  in fetal Down syndrome brain. Biochem Biophys Res Commun 293: 836–841
- Wenstrup RJ, Murad S, Pinnell SR (1989) Ehlers-Danlos syndrome type VI: clinical manifestations of collagen lysyl hydroxylase deficiency. J Pediatr 115: 405–409
- Yeowell HN, Walker LC (2000) Mutations in the lysyl hydroxylase 1 gene that result in enzyme deficiency and the clinical phenotype of Ehlers-Danlos syndrome type VI. Mol Genet Metab 71: 212–224

Authors' address: Prof. Dr. Gert Lubec, CChem, FRSC (UK), Department of Pediatrics, Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria,

Fax:  $+43$  40400 3194, E-mail: gert.lubec@meduniwien.ac.at