Biomarkers, January–February 2005; 10(1): 29–40

Taylor & Francis
Taylor & Francis Group

Identification of stool proteins in C57BL/6J mice by two-dimensional gel electrophoresis and MALDI-TOF mass spectrometry

M. B. OLEKSIEWICZ, H. Ø. KJELDAL, & T. G. KLENØ

Department of Virology and Molecular Toxicology, Novo Nordisk A/S, Novo Nordisk Park, Måløv, Denmark

Abstract
Gastrointestinal disease is a major cause of mortality in humans and animals, and the detection of disease-associated protein in stool is an established diagnostic method in this context. Yet, no data currently exists about the protein composition of mammalian faeces. Using a newly developed two-dimensional (2D) gel method, 28 of the most abundant proteins in murine faeces were identified. Mammalian faeces contains protein from multiple species (from the individual, from gastrointestinal bacteria, from food, etc.). Yet, it was found that the majority of mouse stool proteins were of mouse origin, with a minority of proteins being derived from food (in particular soybean glycinin and conglycinin) and bacteria (flagellin). Most mouse proteins were proteases and saccharidases derived from the exocrine pancreas. In addition, two unexpected mouse proteins were identified: one was a newly described mucin-like protein from intestinal goblet cells (FcγBP); the other was the secreted form of carbonic anhydrase (type VI) from salivary gland. The data suggest that 2D analysis of faecal protein is likely to provide meaningful information about the physiological stage of the gastrointestinal tract. Compared with studies based on biopsies, faecal protein analysis may reduce the number of laboratory animals, and might also allow quicker bridging from animal studies to humans, where biopsy material is more difficult to obtain and is less relevant for general practice use.

Keywords: *Faeces protein, stool protein, two-dimensional gel, biomarker, MALDI-TOF*

(Received 2 November 2004; accepted 17 December 2004)

Introduction

Neoplastic and inflammatory diseases of the gastrointestinal tract are quite important in human medicine. For example, colon cancer is a major cause of cancer mortality in the USA (Levin et al. 2003). Likewise, inflammatory bowel complaints such as celiac disease, ulcerative colitis and Crohn's disease are relatively common in humans (Karlinger et al. 2000).

This heterogenous group of diseases has in common a need for improved diagnosis and screening, preferably by non-invasive means. Stool analysis is the obvious non-invasive technique in this context, and many stool biomarkers already exist. For example, detection of haemoglobin in faeces is a classical method to screen for

Correspondence: M. B. Oleksiewicz, Department of Virology and Molecular Toxicology, Novo Nordisk A/S, Novo Nordisk Park, F9.1.21, DK-2760 Måløv, Denmark. Tel: 45-4442-0074. Fax: 45-4466-3939. Email: mboz@novonordisk.com

ISSN 1354-750X print/ISSN 1366-5804 online
DOI: 10.1080/13547500500038775

colorectal cancer, and detection of calprotectin in stool is being used for the diagnosis of inflammatory bowel complaints (Poullis et al. 2002, Levin et al. 2003). Nevertheless, there is a recognized need for new, better faecal biomarkers for these type of diseases (Poullis et al. 2002).

Often, biomarkers are first identified in animal models, for example by comparing protein and mRNA expression patterns between diseased and healthy animals, by standard high-throughput expression profiling methods such as two-dimensional (2D) gel electrophoresis (protein) or microarray hybridization (mRNA). Typically, biopsies of the affected gastrointestinal tract segment would be used, as, intuitively, biopsies of the affected tissue should provide the most sensitive biomarker detection. Yet, biopsy material has the disadvantage of being an invasive technique, not directly applicable to for example general practice use or large-scale screening in humans. To ensure the quickest possible bridging into human studies, direct examination of faeces would seem an obvious supplementary approach in animal studies, but, surprisingly, no data currently exists about large-scale protein profiling of laboratory animal faeces.

Recently, we described a method for 2D gel analysis of faeces protein from common laboratory animals such as mice, rats and Göttingen minipigs. When mouse faeces is extracted according to our method, between 100 and 200 abundant protein species can be visualized on Coomassie-stained 2D gels (Oleksiewicz 2004). Yet, 2D gel analysis of faecal protein is only relevant if the visualized proteins stem from the mouse's gastrointestinal tract; in biomarker context, bacterial and food proteins, which are obviously also expected in faeces, are probably less relevant. Surprisingly, given the importance of the gastrointestinal tract in human and animal health, no published data exists on the qualitative distribution of proteins in mammalian faeces. Therefore, in the present study, we identified the most abundant faecal proteins of C57BL/6J mice, a strain commonly used in pharmacological research, by 2D gel analysis and MALDI-TOF mass spectrometry.

Materials and methods

Animals

C57BL/6J mice, 14–16 weeks of age, were housed in plain cages bedded with wood chips, in unisex groups of five animals, with acidified water and food (Altromin 1324 rat and mouse maintenance diet; Altromin International, Lage, Germany) ad libitum. The animals did not receive any treatment. The cages were kept in a standard laboratory animal housing unit, under a 12 h light/dark cycle. Faecal pellets that had been deposited overnight were collected from cage bottoms in the morning, taking care to avoid contamination with bedding material, and avoiding soft faecal pellets, which might be contaminated with urine (Oleksiewicz 2004).

Two-dimensional gel analysis and MALDI-TOF mass spectrometric protein identification

Two-dimensional gel analysis of faecal protein was performed as described (Oleksiewicz 2004). Briefly, proteins were first extracted by dissociating faecal pellets in PBS with 10 mM EDTA and protease inhibitors, and then precipitated away from salts and contaminants, before 2D analysis on minigels, which covered the pI range 3–10, and size range of 180 to 6 kDa. Individual protein spots were excised from Coomassie-stained gels and subjected to in-gel trypsin digestion according to standard protocols.

The trypsin digests (20 μl) were lyophilized, reconstituted in 5 μl 5% formic acid, and stored at −20°C.

Part (1/5) of the trypsin digest was used for tryptic peptide mass fingerprinting, essentially as described (Shevchenko et al. 1996, Wilm et al. 1996). A total of 1 μl trypsin digest was mixed with 1 μl matrix (a saturated solution of α-cyano-4-hydroxy-cinnamic acid in 70% acetonitrile, 0.1% trifluoroacetic acid), and 1 μl of this mix was deposited on the MALDI probe (ground steel, Bruker Daltonics, Bremen, Germany) and allowed to crystalize on air. Mass spectra were acquired on a Bruker Ultraflex TOF-TOF mass spectrometer, in reflector mode. Mass spectra were collected in positive-ion mode, using an acceleration voltage of 25 kV and a delay of 150 ns. In order to avoid saturation of the detector by matrix ions, a low-mass detector gating of 500 Da and matrix deflection was used. The instrument was externally calibrated on a mixture of peptides. The resulting tryptic peptide masses were matched at 100 ppm tolerance against the MSDB mammalian database, using the Mascot search engine (http://www.matrixscience.com/home.html).

In cases where tryptic peptide mass fingerprinting did not result in protein identification, or where confirmation of tryptic peptide mass fingerprint data was required, *de novo* peptide sequencing was performed. For this, the remainder (4/5) of the tryptic digest was derivatized with sulfonic acid (Keough et al. 1999, 2000, 2002), using the Ettan CAF (chemically assisted fragmentation) MALDI sequencing kit (Amersham Biosciences, Hørsholm, Denmark). Sulfonic acid modification enhances peptide fragmentation and reduces the complexity of MS/MS spectra to y-ions (Keough et al. 1999, 2000, 2002). Post-source decay spectra were acquired on the Bruker Ultraflex in MS/MS mode. Amino acid sequences were read as differences between y-ion mass/charge (m/z) values (using mass values from http://us.expasy.org/tools/findmod/findmod_masses.html) and protein identification done by BLAST search (http://www.ncbi.nlm.nih.gov/BLAST/).

The following protein identification criteria were used for the data reported in Table I: (1) significant Mascot or BLAST scores ($p < 0.05$); (2) in addition to statistical significance, credible hits were defined as hits against murine, bacterial or common foodstuff (wheat and soy) proteins when searching against unconstrained non-redundant protein database (http://www.ncbi.nlm.nih.gov/BLAST/), or 'all' or 'mammalian' MSDB database subsets (http://www.matrixscience.com/home.html); and (3) the nature of identified murine proteins should be compatible with gastrointestinal tract localization.

Results

In total, we performed four independent experiments, where proteins were extracted from the stool of C57BL/6J mice, and analysed by 2D gel electrophoresis. New mice were used for each experiment, and in each experiment, pooled stool from approximately five animals was used per 2D gel. Each experiment focused on either male (experiment 1) or female (experiments 2 and 3) mice, or included both sexes (experiment 4).

We collected faeces from cage bottoms in the morning, a convenient sampling procedure, but one not allowing control of the interval between defecation and protein extraction. Therefore, protein degradation in deposited faecal samples was an issue. We found that from mice fed a standardized diet, 2D gel patterns were quite

Table I. Proteins identified in murine stool.

[a]Protein spot number	[b]Identity (GenBank accession number)	[c]Predicted molecular weight and pI	[d]Mascot and BLAST *E* values
1	Mouse serum albumin (P07724)	66.7 kDa, pI = 5.6	314 (63%)
2	Mouse IgG Fc binding protein, FcγBP (AAH30871)	not known	157
3	Mouse IgG Fc binding protein, FcγBP (AAH30871)		118
4	Mouse IgG Fc binding protein, FcγBP (AAH30871)		73
5*	Mouse IgG Fc binding protein, FcγBP (AAH30871)		*m/z* 1123 and 2389, nine and 15 amino acid sequence tags, *E* = 4e−10
6	Mouse pancreatic amylase 2 (NP_033799)	55.7 kDa, pI = 7.0	169 (61%)
7	Mouse pancreatic amylase 2 (NP_033799)		72 (16%)
8*	Mouse pancreatic amylase 2 (NP_033799)		*m/z* 1427, 12 amino acid sequence tag, *E* = 7e−04
9	Mouse secreted carbonic anhydrase VI (AAH49973)	34.4 kDa, pI = 5.9	136 (39%)
10*	Mouse pancreatic elastase 3B (NP_080695)	27.3 kDa, pI = 5.3	*m/z* 1330 and 1928, 11 and 12 amino acid sequence tags, *E* = 7e−03
11*	Mouse pancreatic carboxypeptidase B1 (XP_130814)	46.1 kDa, pI = 5.0	*m/z* 1468, 14 amino acid sequence tag, *E* = 1e−03
12*	Mouse pancreatic carboxypeptidase B1 (XP_130814)		*m/z* 1468, 14 amino acid sequence tag, *E* = 1e−03
13	Mouse α-2μ-globulin (CAC34259)	18.7 kDa, pI = 4.6	150 (66%)
14*	Mouse tissue kallikrein type 5 or 6 (A25606 or NP_034769)	27.0 kDa, pI = 4.7	*m/z* 1347 and 1556, 11 and 13 amino acid sequence tags, *E* = 1e−02
15	Mouse mucin MUC 2 (CAD54414)	not known	117
16	Mouse trypsin (B25528)	24.8 kDa, pI = 4.2	97 (28%)
17*	Bacterial flagellin (B32010)	31.4 kDa, pI = 5.4	*m/z* 1132, 1148 (same eight amino acid sequence tag with methionine oxidation) and *m/z* 2296 (21 amino acid sequence tag), *E* = 2e−06
18*	Bacterial flagellin (B32010)	31.4 kDa, pI = 5.4	*m/z* 2296, 21 amino acid sequence tag, *E* = 2e−06
19	Wheat xylanase inhibitor I (CAD19479)	31.2 kDa, pI = 8.5	125 (35%)
20	Soybean trypsin inhibitor (P01070)	24.0 kDa, pI = 4.8	77 (32%)
21	Soybean beta-conglycinin, beta subunit (BAA23361)	48.0 kDa, pI = 5.6	106 (25%)
22	Soybean beta-conglycinin, beta subunit (BAA23361)		88 (25%)
23	Soybean beta-conglycinin, beta subunit (BAA23361)		99 (27%)

Table I (*Continued*)

[a]Protein spot number	[b]Identity (GenBank accession number)	[c]Predicted molecular weight and pI	[d]Mascot and BLAST *E* values
24*	Soybean glycinin, G4, basic chain (AAB23212)	20.6 kDa, pI = 10.1	*m/z* 1502, 13 amino acid sequence tag, *E* = 4e−04
25*	Soybean glycinin, G1, basic chain (S10851)	20.4 kDa, pI = 8.4	*m/z* 1493, 12 amino acid sequence tag, *E* = 9e−03
26*	Soybean glycinin, G1, basic chain (S10851)	20.4 kDa, pI = 8.4	*m/z* 979, 1373, 1426 and 1450, 8–13 amino acid sequence tags, *E* = 6e−03
27*	Soybean glycinin, G1, G2 or G3, basic chain (S10851, CAA33216 and AAB23211)	20.3 to 20.6 kDa, pI = 8.4 (G1 and G2) or 5.6 (G3)	*m/z* 1358, 11 amino acid sequence tag, *E* = 3e−03
28*	Soybean glycinin, G4, basic chain (AAB23212)	20.6 kDa, pI = 10.1	*m/z* 1149, 1374, 1425 and 1570, 10–14 amino acid sequence tags, *E* = 6e−03

[a]Protein spot numbers correspond to 2D gel images in Figure 1. *Protein identification was made by *de novo* peptide sequencing (Keough et al. 1999, 2000, 2002). Spot numbers without asterisks indicate that protein identification was made by tryptic peptide mass fingerprinting. Detailed information about tryptic peptide mass fingerprint spectra and derived *de novo* peptide sequences is available on request.

[b]Proteins are identified by name, with GenBank accession numbers in parentheses. For kallikrein (spot 14), the amino acid sequences derived from *de novo* peptide sequencing could not differentiate between kallikrein types 5 and 6, hence accessions for both are given. For flagellin (spots 17 and 18), the amino acid sequences derived from *de novo* peptide sequencing did not have a perfect GenBank match; the shown GenBank accession is the best homologue, identified by BLAST search. For glycinin (spots 24–28), all the amino acid sequences derived from *de novo* peptide sequencing matched the basic chain, i.e. the carboxyterminal part of the complete glycinin sequence provided by the indicated GenBank accessions (Nielsen et al. 1989, Beilinson et al. 2002). The presence of serum albumin and the basic soy glycinin chain in murine faeces has already been mentioned (Oleksiewicz 2004).

[c]Molecular weights (MW) and isoelectric points (pI) were determined from the indicated GenBank sequences, with DS Gene software (Accelrys, Inc.), excluding signal peptides (identified with SignalP, http://www.cbs.dtu.dk/services/SignalP/#submission). The MW and pI values do not account for post-translational modifications. 'Not known' for MUC2 and FcγBP indicates the sequences provided by the indicated GenBank accession numbers might not be of a full length, according to our own comparison with rat and human sequences (data not shown). For glycinin (spots 24–28), the indicated MW and pI were calculated for the basic chain (carboxyterminal part of the glycinin sequence provided by the GenBank accessions), based on the cleavage sites provided by Nielsen et al. (1989) and Beilinson et al. (2002).

[d]Mascot scores are provided for spots identified by tryptic peptide mass fingerprinting (spot numbers without asterisks). The MSDB database (January 2004) was searched against 'mammalian' entries (spots 1–4, 6, 7, 9, 13, 15, 16) or 'all' entries (spots 19–23), with scores above 67 and 75 being significant, respectively ($p < 0.05$) (http://www.matrixscience.com/). Sequence coverage is provided in parentheses behind the Mascot score, where appropriate. For spots identified by CAF-assisted *de novo* peptide sequencing (spot numbers with asterisks), the *m/z* values of the sequenced peptides are provided as well as and BLAST *E* values for the best sequence tag. All the shown *E* values are significant ($p < 0.05$) using the formula $p = 1 - e^{-E}$ (http://www.ncbi.nlm.nih.gov/BLAST/).

reproducible between experiments (not shown in detail). Furthermore, we found that as regards to the most abundant proteins, faeces from male and female mice produced very similar 2D gel patterns (Figure 1, cf. A and B). Although faeces protein stability was not examined directly, these consistencies between experiments suggested that the protein in deposited faeces was relatively stable, thus indirectly validating the faeces sampling and protein extraction protocols.

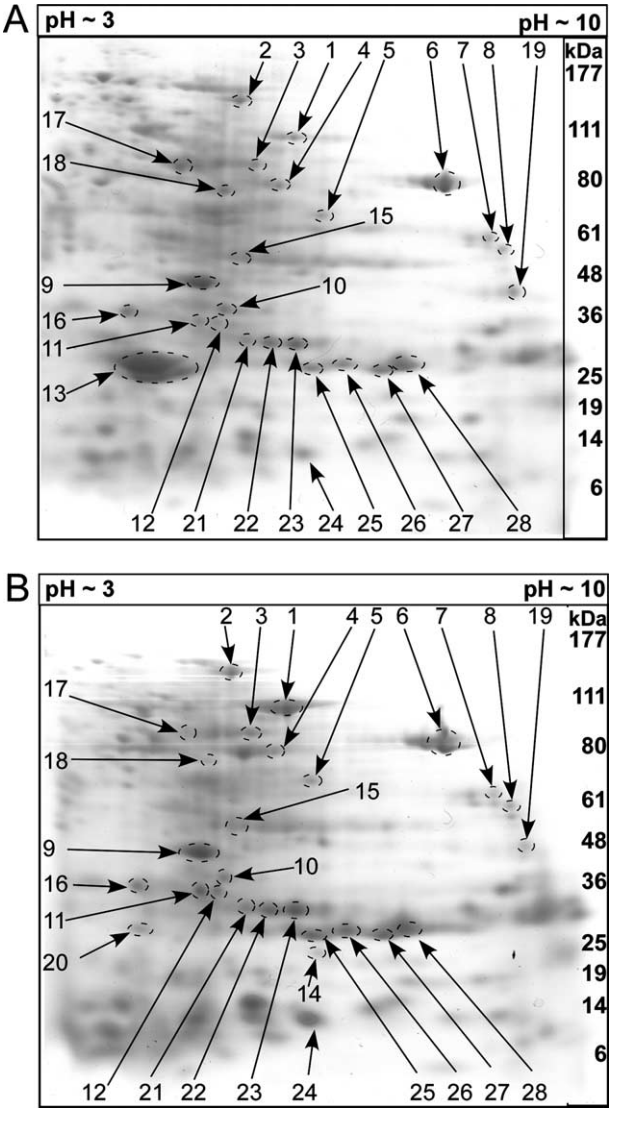

Figure 1. Two-dimensional gel analysis of murine stool protein. A, Male mice; B, female mice. Protein was extracted from pooled faeces, resolved on 2D gels, and Coomassie-stained, as described in the Materials and methods. The indicated pH and molecular weight markers are approximate only. Protein spots that were excised for mass spectrometric protein identification are outlined (dashed-border rings). Discrete protein spots were distinguishable on the original gels, but weak spots can blend with the relatively high, diffuse Coomassie stain background, which seemed to be a feature of the faeces protein gels. The protein spot numbers correspond to those in Table I. Protein spot 14 (kallikrein) on gel B could not be matched on gel A, likely reflecting 2D gel assay variability, although sex differences in kallikrein expression have also been described.

Because it appeared from the above-mentioned results that faeces protein patterns were sufficiently reproducible for 2D analysis to be meaningful, we went on to mass spectrometric protein identification. The faeces protein 2D gels exhibited a slightly higher Coomassie background staining than usually observed in our laboratory for 2D

gels with other material, making graphical reproduction more difficult (Figure 1 and not shown). Nevertheless, in each experiment, at least 100–200 protein spots were easily visualized on the original 2D gels by Coomassie staining (Figure 1 and not shown). We observed that faecal proteins appeared to be quite heterogenous, covering the whole pI and size resolution range of our 2D gel system (Figure 1). Therefore, for identification, we selected protein spots that were abundant (stained strongly with Coomassie), but also attempted to sample proteins with different pI and size values. In each experiment, 48 of the most abundant proteins were excised from the gels according to these criteria, and identified by mass spectrometry. The excised spots largely overlapped between experiments, allowing us to align and match 2D protein spot patterns between all the experiments, and hence compile the mass spectrometric protein identification results from all four experiments in a single gel image (Figure 1, experiment 4). In total, identification was made of 28 spots (Figure 1A and B, identities described in Table I), whereas approximately 20 spots could not be identified, despite in many cases yielding good mass spectra (not shown). As expected, murine proteins, food proteins and bacterial proteins were all found in faeces (Table I). Some proteins were represented by several spots, likely due to post-translational modifications such as glycosylation and proteolytic processing (Figure 1 and Table I).

Interestingly, murine proteins represented the majority of the stool proteins (Table I, 16/28 spots). These abundant murine stool proteins were derived from the salivary glands (Table I, spot 9), the pancreas (Table I, spots 6, 10, 11, 14, 16), serum (Table I, spot 1), and intestinal goblet cells (Table I, spot 15). In terms of spot intensity as well as reproducibility between experiments, murine secreted carbonic anhydrase VI was one of the most abundant faecal proteins (Figure 1 and not shown, spot 9). In terms of spot numbers, a mucin-like protein (the murine homologue of human intestinal Fc-binding protein) was the most abundant, being represented by 4 spots (Figure 1 and Table I, spots 2–5). In some experiments, urinary protein was also observed (Figure 1 and Table I, spot 13). The finding of urinary protein was not constant between experiments (not shown in detail, but for example compare Figure 1A and 1B), and therefore represented occasional contamination of faeces with urine (as opposed to representing ectopic expression of urinary tract protein in the intestine). Because of the quite high abundancy, as well as characteristic size and pI of the major urinary protein (Figure 1A, spot 13), we found that it served as an excellent, objective marker for the quality (presence of urine contamination) of murine faeces samples (Figure 1, compare A, which appeared urine-contaminated, with B, which appeared a urine-free sample).

Discussion

Analysis of faecal protein is an attractive technique for several reasons. First, it allows non-invasive sampling of the gastrointestinal tract environment. Repeated sampling from a single laboratory animal is possible in contrast to mucosal biopsy, which would often involve euthanasia in small laboratory animal species. Thus, faeces analysis might reduce the number of experimental animals required for a study. Second, faecal samples can be taken before and after treatment, allowing the same animal to act as its own control, which would be expected to reduce data variability. Of course, any protocol involving repeated faecal sampling would have to take into account the

gastrointestinal tract passage time. Third, using faecal analysis in laboratory animals may speed up bridging studies to humans, because in humans, faecal analysis is much preferred over mucosal biopsies. The time factor is particularly relevant in biomarker (surrogate endpoint) searches, as biomarkers often have to be identified and validated on a tight time schedule alongside the preclinical development of the experimental drug. Finally, faeces has been implicated in the pathogenesis of Crohn's disease, where it appears that clinical symptoms can be triggered by challenging the colonic mucosa with faeces from Crohn's patients (Harper et al. 1985, Winslet et al. 1994). Presumably, faeces from Crohn's patients may contain disease-aggravating factor(s) (Harper et al. 1985, Winslet et al. 1994), the identification of which is of course relevant, as the aetiology of Crohn's disease is not known, and generally effective medical treatment does not exist.

Despite the relevance of faeces protein analysis outlined above, the methods available are few. In fact, without exception, all published methods appear to target single faecal proteins chosen based on previous knowledge of disease pathogenesis (Poullis et al. 2002, Levin et al. 2003). For example, neutrophil granulocytes are known to infiltrate inflamed intestinal mucosa, and the cytosol of neutrophil granulocytes is known to be very rich in calprotectin, a soluble protein binding calcium and zinc. Based on this knowledge, ELISAs have been developed to quantitate stool calprotectin, and this analysis has in fact shown value for diagnosis and management of inflammatory bowel disease (Poullis et al. 2002). Yet, when identifying biomarkers for the action of new drugs, or exploring the mechanism of action of new drugs, or exploring the pathogenesis of diseases with unknown aetiology, the obvious lack of up-front mechanistic knowledge dictates that evaluation of large numbers of faecal proteins would be required, to identify those that are relevant for the given disease or drug.

The obvious solution to the requirement for 'naïve' high-throughput screening of faecal proteins is 'proteomics'-type approaches, and Oleksiewicz (2004) reported an extraction and purification protocol that allows faecal proteins to be separated on 2D PAGE gels. Such a method has application in, for example, drug development, as it allows the identification of stool proteins that are differentially expressed in drug-treated animals; such drug-regulated proteins might serve as biomarkers (surrogate endpoints) for treatment, or might provide information about drug effect mechanisms. Yet, for this analysis to be meaningful requires that a significant number of the proteins that can be visualized on the 2D gels are of host (in this case, murine) origin, as host proteins would be most relevant for biomarker studies. Intuitively, odds would appear to be against host proteins forming a significant percentage of stool protein, because as a matter of course, the gastrointestinal tract contains a large number of food and bacterial proteins. If bacterial and food proteins obscure host proteins, 2D gel analysis of faecal protein would not add value to biomarker or drug mechanism studies. Surprisingly, despite the importance of gastrointestinal tract disease in human and veterinary medicine, we did not find any published information about the qualitative protein composition of mammalian faeces. Therefore, in the present study, using our new 2D gel method (Oleksiewicz 2004), we determined the origin of the most abundant stool proteins of normal, healthy, untreated mice by mass spectrometry.

Currently, in mass spectrometric protein identification, a bias will exist towards identifying proteins from species for which a high degree of sequence coverage exists.

While the complete murine genome sequence is known, the database coverage for bacteria and common food constituents (e.g. the main proteins of soybeans and wheat grain) is also quite good, and we therefore hoped that bias in protein identification would be minimal. To reduce identification bias further, two different protein identification methods were used: tryptic peptide mass fingerprinting and *de novo* peptide sequencing. We were gratified to observe that approximately 60% of the most abundant faecal proteins could be identified (28 identified out of approximately 48 examined in total). In our experience, even when examining only protein from 'fully sequenced' species, at most around 80–90% of the protein spots excised from 2D gels would yield identification in a first pass effort. Therefore, we believe that meaningful conclusions can be drawn from the species distribution of the 28 stool proteins that were identified (Table I). Thus, we consider it a major, and slightly surprising finding that a majority of the stool proteins were of host (mouse) origin. Excepting α-2μ-globulin from the analysis, as this protein likely represented urine contamination of faeces, 15/27 faecal protein spots, i.e. 56%, represented murine proteins (Figure 1 and Table I, spot 13 excluded from the analysis). Some proteins were found in several spots, most likely due to post-translational processing, such as glycosylation and proteolysis (Figure 1 and Table I, e.g. spots 2–5 represented the same protein). Therefore, the 15 murine spots represented nine mouse stool proteins (Table I, again excepting spot 13). As expected, bacterial and food (plant) proteins were also identified (Table I, spots 17–28).

Another interesting finding in the present study was the tissue origin and function of the host (mouse) stool proteins: five proteins were proteases and saccharidases derived from the exocrine pancreas (Figure 1 and Table I, spots 6–8, 10, 11, 14, 16). This was not surprising, because proteases and saccharidases are expected to be abundant as well as stable in the gastrointestinal tract, and the exocrine pancreas is known as a rich source of such enzymes. Two proteins were derived from the intestinal mucosa (Figure 1 and Table I, spots 2–5, 15). MUC2 (Figure 1 and Table I, spot 15) is a mucin secreted from colonic goblet cells. MUC2 is present in the protective gel layer of the colonic mucosa in a glycosylated and polymerized, very high molecular weight form ($>10^7$). However, as part of the normal mucin turnover, MUC2 is known to be proteolytically processed into smaller fragments, that are highly resistant to further breakdown (Allen et al. 1998). Therefore, the presence of a MUC2 fragment in faeces was not surprising (Figure 1 and Table I, spot 15). Spots 2–5 represented the murine homologue of a recently cloned human IgGFc binding protein (FcγBP, Table I). FcγBP is produced mainly in goblet cell in the large and small intestine, and has sequence homology to mucins (Harada et al. 1997, Kobayashi et al. 2002). Additionally, human FcγBP has an immunoglobulin Fc-binding activity that is absent in mucins. Therefore, it is assumed to be involved not only in mucin-like mechanistic protection of the intestinal mucosal surface, but also in regulating mucosal immunity and intestinal inflammation (Harada et al. 1997, Kobayashi et al. 2002). Furthermore, FcγBP was recently shown to be upregulated in the adapting rat intestine, and is also regulated during embryonal development (Rubin et al. 2000). Based on the above, the presence of MUC2 and FcγBP in faeces was not surprising. However, their relative abundance was quite unexpected: MUC2 has been described as being the predominant secretory colonic mucin (Allen et al. 1998). In contrast, our results indicate that at least in the faecal stream, FcγBP is at least as abundant (Figure 1 and Table I). Our study provides the first information about the murine FcγBP homologue

(this protein has so far only been described in rats and humans), and shows that this protein is unexpectedly abundant in the faecal stream. This supports that the newly discovered FcγBP may have important physiological roles.

In similar vein, we were surprised to find that one of the most abundant faecal proteins was of salivary gland provenance (Figure 1 and Table I, spot 9). The spectra we derived from spot 9 identified this protein unambiguously as carbonic anhydrase type VI, the only secreted form of the mammalian carboanhydrase family, which is exclusively expressed in the parotid and submandibular salivary glands (Parkkila et al. 1994, Kivela et al. 1999). We have found that the carbonic anhydrase VI spot is highly reproducible on 2D gels of murine faecal protein (e.g. easily identifiable on previously published gels from our laboratory) (Oleksiewicz 2004). Carbonic anhydrase VI is present at 5–30 mg ml^{-1} in saliva (about 3% of total salivary protein), and provides the main buffering capacity of saliva, a buffering capacity which is crucial for dental health (Parkkila et al. 1994, Kivela et al. 1999). Remarkably, carbonic anhydrase VI has been suggested to retain activity in the gastric lumen, and function in the protection of the oesophageal, gastric and duodenal mucosa against acid injury (Parkkila et al. 1997). While we do not know whether spot 9 (Figure 1) represents active enzyme, our findings nevertheless support that carbonic anhydrase VI is quite stable in the gastrointestinal tract, and may have functions downstream of the oral cavity.

Proteolysis is an important issue when analysing stool protein because obviously, the gastrointestinal tract is rich in proteolytic enzymes. Further, we believe that not only the frequency of host protein, but also the integrity of host protein is of importance for the utility of stool protein analysis, as the biological information gained from detection of degradation products might be lesser than from detection of full-length (functional) proteins. Two observations suggest that at least some of the abundant host proteins were full-length (not proteolytically degraded) molecules. First, for mouse faecal proteins such as albumin, carbonic anhydrase VI, elastase 3B, carboxypeptidase B1, kallikrein and trypsin, we observed only one spot, that migrated in the second dimension PAGE gels close to, or in some cases slower, than predicted from the amino acid sequence (compare predicted MW in Table I with migration in Figure 1). Second, of the nine mouse faecal proteins (Table I, spots 1–12, 14–16), only three were represented by several spots (Figure 1 and Table I, spots 2, 6). Thus, the size and charge heterogeneity which would be expected to result from proteolytic degradation was only observed for a minority of murine faecal proteins. Yet, it should be mentioned that proteolysis is not only involved in protein degradation and inactivation, but often in post-translational regulation of protein function. Therefore, we were intrigued to observe that FcγBP exhibited extensive size and charge heterogeneity on 2D gels (Figure 1, spots 2–5). It is very likely that at least part of this heterogeneity was caused by proteolysis, and it is an intriguing possibility that such processing may be involved in the so far unknown functions of FcγBP.

While our focus was on mouse (host) proteins, we note that bacterial as well as food proteins were also easily identified (Table I and Figure 1, spots 17–28). We have previously described finding the basic chain of soybean glycinin in mouse faeces (Oleksiewicz 2004), and the present study found that the basic chain of soybean glycinin was in fact the most abundant plant protein in laboratory mice stool (Figure 1 and Table I, spots 24–28). This was in agreement with soy being a common constituent of laboratory rodent chow, glycinin being a major soybean storage protein

(Nielsen et al. 1989, Beilinson et al. 2002), and the basic glycinin chain being particularly resistant to digestion in the mammalian gastrointestinal tract (Lalles et al. 1999, Brandon & Friedman 2002). The series of spots formed by the basic soy glycinin chain was highly characteristic and reproducible between experiments (Figure 1 and Table I, spots 24–28). The presence of a series of spots, all containing the basic chain of glycinin (Figure 1 and Table I, spots 24–28), could be due to different glycinin types in the food (Table I), or chemical modifications of glycinin during the heat treatment that is used as part of soybean food processing (Friedman & Brandon 2001, Brandon & Friedman 2002). Also, the series of basic chain glycinin spots could reflect post-translational modifications during gastrointestinal tract passage. The ability of our 2D gel method to monitor the breakdown of glycinin in the gastrointestinal tract may be of relevance in examining the pathogenesis of soybean food allergy, or developing transgenic plants for improved digestibility (Sicherer et al. 2000, Friedman & Brandon 2001). The ability to identify bacterial protein, on the other hand, may be helpful in characterizing the gastrointestinal bacterial flora.

In summary, we performed the first analysis of the qualitative species composition of mammalian faeces protein. Our findings provide the first evidence that proteomic-type analysis of faecal protein may provide biologically meaningful information about the function of the gastrointestinal tract. Our laboratory is engaged in biomarker identification in drug development, and we found it intriguing that several of the faecal proteins we identified have in fact already been suggested as biomarkers (Dominici & Franzini 2002, O'Donovan et al. 2002, Moridani & Bromberg 2003). This may be a serendipitous characteristic of 2D faecal protein analysis: because the protein subset visible on 2D gels has been proteolytically selected by passage through the gastrointestinal tract, it may perhaps be enriched for proteins and/or protein subunits that have a functional role on the mucosal surface and in the gastrointestinal tract lumen.

Acknowledgements

Hanne Toftelund, Anja Bordig and Peter Thygesen are thanked for providing mouse faeces samples, Peter Højrup is thanked for help with mass spectrometry, and Torben Storgaard is thanked for a critical reading of the manuscript.

References

Allen A, Hutton DA, Pearson JP. 1998. The MUC2 gene product: a human intestinal mucin. International Journal of Biochemistry and Cell Biology 30:797–801.

Beilinson V, Chen Z, Shoemaker C, Fischer L, Goldberg B, Nielsen C. 2002. Genomic organization of glycinin genes in soybean. Theory in Applied Genetics 104:1132–1140.

Brandon DL, Friedman M. 2002. Immunoassays of soy proteins. Journal of Agriculture and Food Chemistry 50:6635–6642.

Dominici R, Franzini C. 2002. Faecal elastase-1 as a test for pancreatic function: a review. Clinical Chemistry and Laboratory Medicine 40:325–332.

Friedman M, Brandon DL. 2001. Nutritional and health benefits of soy proteins. Journal of Agriculture and Food Chemistry 49:1069–1086.

Harada N, Iijima S, Kobayashi K, Yoshida T, Brown WR, Hibi T, Oshima A, Morikawa M. 1997. Human IgGFc binding protein (FcgammaBP) in colonic epithelial cells exhibits mucin-like structure. Journal of Biological Chemistry 272:15232–15241.

Harper PH, Lee EC, Kettlewell MG, Bennett MK, Jewell DP. 1985. Role of the faecal stream in the maintenance of Crohn's colitis. Gut 26:279–284.

Karlinger K, Gyorke T, Mako E, Mester A, Tarjan Z. 2000. The epidemiology and the pathogenesis of inflammatory bowel disease. European Journal of Radiology 35:154–167.

Keough T, Lacey MP, Youngquist RS. 2000. Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry. Rapid Communications in Mass Spectrometry 14:2348–2356.

Keough T, Lacey MP, Youngquist RS. 2002. Solid-phase derivatization of tryptic peptides for rapid protein identification by matrix-assisted laser desorption/ionization mass spectrometry. Rapid Communications in Mass Spectrometry 16:1003–1015.

Keough T, Youngquist RS, Lacey MP. 1999. A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry. Proceedings of the National Academy of Sciences. USA 96:7131–7136.

Kivela J, Parkkila S, Parkkila AK, Leinonen J, Rajaniemi H. 1999. Salivary carbonic anhydrase isoenzyme VI. Journal of Physiology 520:315–320.

Kobayashi K, Ogata H, Morikawa M, Iijima S, Harada N, Yoshida T, Brown WR, Inoue N, Hamada Y, Ishii H, Watanabe M, Hibi T. 2002. Distribution and partial characterisation of IgG Fc binding protein in various mucin producing cells and body fluids. Gut 51:169–176.

Lalles JP, Tukur HM, Salgado P, Mills EN, Morgan MR, Quillien L, Levieux D, Toullec R. 1999. Immunochemical studies on gastric and intestinal digestion of soybean glycinin and beta-conglycinin in vivo. Journal of Agriculture and Food Chemistry 47:2797–2806.

Levin B, Brooks D, Smith RA, Stone A. 2003. Emerging technologies in screening for colorectal cancer: CT colonography, immunochemical fecal occult blood tests, and stool screening using molecular markers. CAA Cancer Journal for Clinicians 53:44–55.

Moridani MY, Bromberg IL. 2003. Lipase and pancreatic amylase versus total amylase as biomarkers of pancreatitis: an analytical investigation. Clinical Biochemistry 36:31–33.

Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallon BJ, Fischer RL, Sims TL, Drews GN, Goldberg RB. 1989. Characterization of the glycinin gene family in soybean. Plant Cell 1:313–328.

O'Donovan M, Fischer A, Abdo EM, Simon F, Peter HJ, Gerber H, Buergi U, Marti U. 2002. Differential expression of IgG Fc binding protein (FcgammaBP) in human normal thyroid tissue, thyroid adenomas and thyroid carcinomas. Journal of Endocrinology 174:517–524.

Oleksiewicz MB. 2004. Stool protein analysis by 2D gel electrophoresis. Scandinavian Journal of Gastroenterology 39:787–788.

Parkkila S, Parkkila AK, Juvonen T, Rajaniemi H. 1994. Distribution of the carbonic anhydrase isoenzymes I, II, and VI in the human alimentary tract. Gut 35:646–650.

Parkkila S, Parkkila AK, Lehtola J, Reinila A, Sodervik HJ, Rannisto M, Rajaniemi H. 1997. Salivary carbonic anhydrase protects gastroesophageal mucosa from acid injury. Digestion Diseases Science 42:1013–1019.

Poullis A, Foster R, Northfield TC, Mendall MA. 2002. Review article: faecal markers in the assessment of activity in inflammatory bowel disease. Alimentary Pharmacology and Therapy 16:675–681.

Rubin DC, Swietlicki EA, Iordanov H, Fritsch C, Levin MS. 2000. Novel goblet cell gene related to IgGFcgammaBP is regulated in adapting gut after small bowel resection. American Journal of Physiology of the Gastrointestine and Liver Physiology 279:G1003–G1010.

Shevchenko A, Wilm M, Vorm O, Mann M. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Annals of Chemistry 68:850–858.

Sicherer SH, Sampson HA, Burks AW. 2000. Peanut and soy allergy: a clinical and therapeutic dilemma. Allergy 55:515–521.

Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T, Mann M. 1996. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature 379:466–469.

Winslet MC, Allan A, Poxon V, Youngs D, Keighley MR. 1994. Faecal diversion for Crohn's colitis: a model to study the role of the faecal stream in the inflammatory process. Gut 35:236–242.