IS METABOLOMICS REACHABLE?. DIFFERENT PURIFICATION STRATEGIES OF HUMAN COLON CANCER CELLS PROVIDE DIFFERENT CE-MS METABOLITE PROFILES

Carolina Simó[1], *Clara Ibáñez*[1], *Ángeles Gómez-Martínez*[2], *José A. Ferragut*[2], *Alejandro Cifuentes*[1,*]

[1]Institute of Food Science Research (CIAL), CSIC, Nicolas Cabrera 9, Campus de Cantoblanco, 28049 Madrid, Spain.

[2]Institute of Molecular and Cellular Biology, Miguel Hernandez University, Avda. Universidad s/n; 03202 Elche, Alicante, Spain.

Abbreviations: (AEBSF) 4-(2-aminoethyl)benzenesulfonyl fluoride, (BPE) base peak electropherogram, (EIE) extracted ion electropherogram, (GABA) γ-aminobutyric acid, (PBS) phosphate buffered saline, (PIPES) 1,4-piperazinediethanesulfonic acid.

Running title: Is Metabolomics reachable?. CE-MS of colon cancer metabolites.

Keywords: CE-MS, HT29 cells, Foodomics, Metabolites, Metabolomics, Purification strategies.

Total number of words including figure and table legends: 5493

***Corresponding author:** a.cifuentes@csic.es

ABSTRACT

In this work, four different metabolite purification approaches are investigated prior to Metabolomics of human HT29 colon cancer cells. Namely, methanol deproteinization, ultrafiltration and two solid phase extraction (SPE) methods using C18 and polymer-based cartridges were studied. The extracts were characterized via a metabolomic approach based on the application of capillary electrophoresis time-of-flight mass spectrometry (CE-MS). CE-MS analysis time was less than 20 min per sample and allowed the simultaneous and reproducible analysis of more than 80 metabolites in a single run with a minimum consumption of sample and reagents. Metabolome analysis revealed in some cases important differences among the studied metabolite purification procedures. No significant differences were observed in the metabolite profile using C18 and polymer-based cartridges, or between ultrafiltration and methanol deproteinization. However, important differences were observed in the metabolomic profiles obtained from SPE and methanol deproteinization samples. These results demonstrate the crucial role of the metabolite purification strategy in Metabolomics since it can bias (and in some cases mislead) the conclusions achieved by the metabolomic study.

1. INTRODUCTION

The general aim of any metabolomic study applied to a living cell, tissue or biofluid is to gain detailed insight of the molecular mechanisms governing the metabolic pathways. Like proteome, metabolome is dynamic and highly variable among cell types, organisms, individuals, environments, etc. Thus, one of the main challenges in Metabolomics is to overcome the chemical complexity, heterogeneity and wide dynamic concentration range of endogenous metabolites (amino acids, amines, small peptides, nucleic acids, carbohydrates, organic acids, vitamins, steroids, coenzymes, etc.) present in a biological sample. Metabolomics presents also unique challenges for separation and detection techniques. No single analytical methodology or platform is available to detect, quantify and identify all metabolites in a certain sample. Two analytical platforms are currently used for metabolomic analyses: MS and NMR-based systems. NMR, that in some cases does not require previous analyte separation (and requires minimal sample pre-cleaning), provides detailed information on the molecular structure of compounds complementary to MS-based metabolomic data, although at the expense of low sensitivity. Improved mass spectrometers with better sensitivity and superior mass accuracy and resolution are aimed to the identification and quantitation of complex metabolite mixtures in a single experiment. The use of high and ultra-high resolution mass analyzers (TOF, FTMS, Orbitrap®, etc) is essential to obtain accurate mass measurements for the determination of elemental compositions of metabolites and to carry out tentative identification based on metabolites databases. On the other hand, MS/MS or MS^n experiments, especially when product ions are accurately analyzed at high resolution (namely, Q-TOF, TOF-TOF, LTQ-Orbitrap®), provide useful additional structural information for the identification of the metabolites. These techniques either

standalone or combined with separation techniques (typically, LC-NMR, GC-MS, LC-MS and CE-MS), are capable of producing complementary analytical information to get a more extensive metabolome coverage [1-5]. Systematic profiling/fingerprinting of as many metabolites as possible has gained broad interest during the last decade. CE ~~technique~~ is particularly suited for the rapid separation of ionic and polar compounds with very high resolution using extremely small reagents and sample volumes. Moreover, no pre-column derivatization of analytes is necessary. On the contrary, lower sensitivity and higher variability of migration times are generally obtained compared to LC or GC. TOF mass analyzer is preferably used in CE-MS due to its ~~high scan speed~~ high spectral acquisition rate and high mass resolution, what perfectly fits with the narrow peaks provided by CE. The capacity of CE-MS to analyze complex mixtures of metabolites in short times opens interesting possibilities in the growing Metabolomic area. Until today, CE-MS has successfully been applied in many metabolomic studies. Interesting reviews have already been published describing the potential of CE-MS in Metabolomics [6-9].

In general, non-targeted metabolomics is addressed to detect as many metabolites as possible in a certain sample. However, at present there is not a general applicable sample preparation protocol to extract the whole range of endogenous metabolites present in a biological sample. Sample preparation is a critical step in any analytical method with important consequences in the final results. In order to prevent loss of metabolites in non-targeted Metabolomics minimal sample treatment should be carried out before analysis. The procedure used for metabolite extraction has to be robust and highly reproducible. It will depend on both the sample type and the targeted metabolites of interest (fingerprinting or profiling approach). Most sample extraction techniques are more or less selective, thus the choice of the appropriate one is very critical for

metabolomic studies since it can bias the final results obtained. In a non-targeted metabolomic work there is a clear need for the development of methods that enable a comprehensive characterization of the metabolome. In this work, CE-MS was used to study the metabolome of human HT29 colon cancer cells. For non-targeted Metabolomics of biological samples deproteinization with an organic solvent is often carried out to avoid adsorption of proteins to the inner capillary wall and the consequent damage on metabolite separation and ion suppression at the ion source. In order to minimize sample handling and avoid extra variability to the final statistic study required in any non-targeted metabolomic study, several common extraction procedures (protein precipitation, SPE, filtration) were selected among the enormity of possibilities due to their simplicity, effectiveness and recognized reproducibility. Namely, four metabolite purification approaches were systematically compared in this work (i.e., two different SPE methods, protein precipitation with methanol and ultrafiltration). The metabolomic profiles obtained were compared based on the total number and type of extracted metabolites, using the information provided by CE-MS for their tentative identification. The results give an additional prove on the difficulty to achieve a representative metabolite profile in Metabolomics.

2. MATERIALS AND METHODS

2.1. Reagents

All chemicals were of analytical reagent grade and used as received. Methanol used in the metabolite extraction procedure was from Sigma-Aldrich (St. Louis, MO). All reagents and solvents employed in the preparation of CE electrolytes and sheath liquids

were of MS grade: formic acid and 2-propanol were from Riedel-de Haën (Seelze, Germany), and water was from Scharlau (Barcelona, Spain). Amino acids from Sigma-Aldrich were dissolved in purified water deionized by using a Milli-Q system from Millipore (Bedford, MA, USA), at the following concentrations: 0.58 mg/mL arginine, 0.49 mg/mL lysine, 0.52 mg/mL histidine, 0.34 mg/mL γ-aminobutyric acid (GABA), 0.39 mg/mL valine, 0.35 mg/mL serine, 0.44 mg/mL leucine, 0.40 mg/mL threonine, 0.49 mg/mL glutamine, 0.38 mg/mL proline, 0.44 mg/mL aspartic acid and 0.09 mg/mL tyrosine. Tyramine, DL-methionine sulfone and 1,4-piperazinediethanesulfonic acid (PIPES) from Sigma-Aldrich were selected as internal standards.

2.2. Samples

Human colorectal adenocarcinoma HT29 cells were used in all the experiments. HT29 cells were grown in DMEM supplemented with 5% heat-inactivated fetal calf serum, 2 mM of L-glutamine, 50 U/ml of penicillin G and 50 µg/ml of streptomycin, at 37 ºC in humidified atmosphere and 5 % CO_2.

For sample preparation, a phosphate buffered saline (PBS) solution containing 138 mM sodium chloride, 2.7 mM potassium chloride and 10 mM sodium hydrogen phosphate, at pH 7.4, was purchased from Sigma-Aldrich. Composition of homogenization buffer was next: 10 mM Tris-HCl, 5 mM EDTA, 120 mM NaCl, at pH 7.4, all of them from Sigma-Aldrich. A protease inhibitor cocktail containing 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), pepstatin A, E-64, bestatin, leupeptin, and aprotinin, was purchased from Sigma-Aldrich.

Human HT29 colon cancer cells were washed with PBS solution and centrifuged. The pellet was resuspended with homogenization buffer and protease inhibitor cocktail. Cells were disrupted with a Polytron homogenizer and centrifuged (14 min at 14000 g and 4ºC).

Pellet (nuclear fraction) was discarded and supernatant was centrifuged for 1h at 100000 g and 4ºC. Supernatant (cytosolic fraction) was stored at -80º C until metabolite purification procedure was carried out. At this temperature, enzyme activity is stopped and samples can safely be stored without continuing metabolic activity. The total protein content was determined by the Bradford method using a commercial dye reagent from Bio-Rad (Hercules, CA, USA) and using BSA as standard.

2.3. Metabolite extraction procedures

Four different metabolite extraction procedures were studied in this work: two SPE methods using different sorbents, protein precipitation with methanol and ultrafiltration.

2.3.1. Solid-phase extraction (SPE)

Two different sorbents were investigated, namely, Isolute C18 Endcapped cartridges (100 mg) from Biotage (Cardiff, Wales, UK), and Evolute^{TM} ABN columns (25 mg). Lower sorbent mass was selected for the last as a result of the higher capacity of the polymeric sorbents due to their higher specific surface area. In both cases, activation, conditioning, and elution were carried out in identical conditions. Activation and conditioning were carried out by passing 1 mL of methanol followed by 1 mL of water through the cartridges. 350 µL of water were added to 150 µL of cytosolic fraction

obtained from the cell culture. This solution was loaded onto the column. After sample loading sorbent was washed with 1 mL of water-methanol (95:5, v/v). Sample elution was performed with 500 µL of methanol. Eluted sample was then aliquot in 100 µL volume.

2.3.2. Protein precipitation

For protein precipitation, 350 µL of methanol were added to 150 µL of cytosolic fraction sample obtained from the cell culture. The solution was incubated at -20ºC during 2h. After incubation the suspension was centrifuged at 20000 *g* and 4ºC for 5 min. The pellet was discarded and the supernatant fraction was collected and aliquot in 100 µL volume.

2.3.3. Ultracentrifugation

350 µL of water were added to 150 µL of cytosolic fraction sample obtained from the cell culture. This solution was ultrafiltrated using an Amicon Ultra 3 kDa centrifugal device (70 min at 14000 g and 4ºC) from Millipore (Bedford, MA, USA). Fraction with molecular weight lower than 3 kDa was collected and aliquot in 100 µL volume.

In all cases, after metabolite purification procedures, the obtained 100 µL aliquots were vacuum-dried. The dried extracts were stored at -80ºC until used. Prior to CE-MS analysis, dried extracts were dissolved in 20 µL of water of MS grade from Scharlau.

2.4. Instrumentation

CE analyses were carried out in a P/ACE 5500 CE apparatus from Beckman Instruments (Fullerton, CA, USA). The instrument was controlled by a PC running the System Gold software from Beckman. Uncoated fused-silica capillaries (50 μm id and 90 cm total length) from Composite Metal Services (Worcester, England) were coupled to MS through an orthogonal ESI interface model G1607A from Agilent Technologies (Palo Alto, CA, USA). Electrical contact at the electrospray needle tip was established via a sheath liquid. A TOF MS instrument (micrOTOF) from Bruker Daltonics (Bremen, Germany) was employed. The instrument was controlled by a PC running the micrOTOF control software from Bruker Daltonics.

2.5. CE-ESI-TOF MS conditions

Before first use, the separation capillary was conditioned by rinsing with 1 M NaOH for 10 min, followed by 20 min with Milli-Q water and 5 min with the separation buffer. After each run, the capillary was conditioned with Milli-Q water during 4 min, followed by separation buffer during 4 min. Injections (53 nL of sample volume) were made at the anodic end using N_2 pressure of 0.5 psi (34.5 mbar) for 80 s. The electrophoretic separation was achieved using +25 kV as running voltage at a constant temperature of 25 ºC in a 1M formic acid BGE. Electrical contact at the electrospray needle tip was established via a sheath liquid based on isopropanol-water (50:50, v/v) and delivered at a flow rate of 0.24 mL/min by a 74900-00-05 Cole Palmer syringe pump (Vernon Hills, IL, USA). The mass spectrometer operated with the ESI source in the positive ion mode. The nebulizer and drying gas conditions were 0.4 bar N_2 and 4 L/min N_2, respectively, and maintaining the ESI chamber at 200ºC. The micrOTOF was operated to acquire spectra in the m/z range of 50-600 every 90 ms. The accurate mass data of the

molecular ions were processed using the DataAnalysis 3.3 software from Bruker Daltonics. External and internal calibration of the TOF MS instrument was performed by introducing a 5 mM sodium formate solution through the separation capillary. Masses for the calibration of the TOF MS instrument were next: 90.9766, 158.9641, 226.9515, 294.9389, 362.9263, 430.9138, 498.9012 and 566.8886 m/z. Each sample was analyzed in triplicate by CE-MS.

2.6. Data processing

Calculation of the elemental composition of compounds was carried out using the Generate Molecular Formula Editor within DataAnalysis software from Bruker Daltonics. Accurate m/z value and migration time from each peak was annotated. Redundant responses from the same ion, such as isotopic peaks, fragments, adducts, dimers, etc, were removed based on established m/z differences. Spike noise and low reliability signals (no peak-like shape) were also eliminated. For the calculation of the total number of the different metabolites detected by CE-MS after each extraction protocol, only those metabolites that repeatedly appeared in three consecutive replicates were considered. TOF MS provided a high mass resolution and high mass accuracy with errors usually below 5-10 ppm. Selected mass spectra were processed through the software DataAnalysis, which provided a list of possible elemental formulas by using the Generate-MolecularFormula Editor, which provided standard functionalities such as minimum/maximum elemental range, electron configuration and ring-plus double bonds equivalents, as well as a comparison between the theoretical and the experimental isotopic pattern (Sigma-Value^{TM}) for increased confidence in the theoretical molecular formula. Tentative identification based on the obtained theoretical molecular formula

was carried out with different free available databases: Human Metabolome Database (HMDB) [10], Metlin [11], KEGG compound [12-14], and PubChem (http://pubchem.ncbi.nlm.nih.gov/). In cases where databases offered more than one possible metabolite for one molecular formula, migration time provided by electrophoretic separation was studied to elucidate the expected electrophoretic mobility of that compound at the separation pH. When available, standards were used to confirm metabolite identification. Metabolite relative levels were calculated from CE-MS data using peak areas. Freely MassTRIX web server [15] was used to map the identified metabolites to KEGG pathways.

3. RESULTS AND DISCUSSION

3.1. Sample preparation

Considering the great diversity and heterogeneity of metabolites, simultaneous purification and analysis of all metabolites from a biological sample is still a challenge. Sample preparation (extraction, clean-up, concentration, etc.) for metabolomic studies depends on the type of sample being analyzed, the subsequent method of analysis and the goal of the metabolic work (whether targeted or non-targeted metabolomic study is going to be carried out). In many cases, sample must be pre-treated in order to eliminate interfering matrix constituents which can negatively affect the metabolomic results.

In CE-MS, complex biological samples containing components that interact with the inner silica capillary surface, lead to alterations in EOF and subsequent lack of reproducibility between injections. More precisely, one of the main problems when

working with biological samples is the adsorption of proteins onto the capillary wall that may produce changes on EOF. On the other hand, high-salt containing sample could reduce efficiency due to electromigration dispersion, producing irreproducible injections (in case of electrokinetic injection) and migration times. Matrix effects can also lead to difficulties in the detection of certain compounds by MS since ESI-MS detection is sensitive to the presence e.g., of salts or other compounds that can comigrate with the analytes of interest.

In this work, a lysate from human HT29 colon cancer cells with rather high salt content (see "Samples" section) and a total protein concentration of 0.9 mg/mL was analyzed using different sample preparation methodologies. Two different procedures were assayed to remove proteins from these samples: protein precipitation with methanol and deproteinization by ultrafiltration using a membrane pore size of 3 kDa. Using any of the mentioned procedures, only deproteinization is carried out, while salts remained in the final extract. On the other hand, offline SPE was also assayed for both sample deproteinization and desalting of samples, while additionally it can enrich the more retained analytes. A wide range of chemically modified sorbent materials (silica gel or synthetic resins, modified or not) enable metabolite purification based on different types of physicochemical interactions. Reserved-phase was selected as it is widely used in many applications included Metabolomics. The use of polymeric sorbents is gaining more attention as a method for metabolite purification in non-targeted approaches. Thus, two different SPE sorbents were studied in this work: a classical octadecyl bonded endcapped silica (C18) sorbent and a polymer-based sorbent (ABN). It is expected that the C18 SPE retains compounds of mid to low polarity due to their polar interaction with the sorbent. On the other hand, ABN uses a polystyrene-divinylbenzene

(PS-DVB) sorbent, suitable for the extraction of a wide range of analytes (acidic, basic and neutral). The same extraction protocol was selected for both types of sorbents. In SPE loss of metabolites during sample loading and washing steps is generally unavoidable. In our case, washing fractions can contain small hydrophilic and charged metabolites that will be lost. However, together with these potentially interesting metabolites, interfering compounds as salts and proteins from the cytoplasm are expected to be also present. We decided to discard this fraction in order to avoid protein adsorption onto the inner capillary wall what could affect migration time reproducibility and metabolite signal suppression at the ESI ion source. The same cytosol sample was subjected to these four purification procedures. After metabolite purification, extracts were directly analyzed by CE-MS.

3.2. CE-ESI-TOF MS method development

A low pH was selected for CE-MS analysis in order to both avoid analyte adsorption onto the inner capillary wall and confer positive charge to the analytes to improve their ESI ionization yield. A BGE composed of 1 M formic acid in water at pH 1.8 was used. At the low pH electrolyte used in this work most amino acids, amines, nucleosides, small peptides, and in general, basic compounds, present overall positive charge and they migrate before EOF. Using these analytical conditions, CE-MS method is focused on cationic metabolites. However, using these electrophoretic conditions it was also possible to detect some acidic compounds (bearing simultaneously negative and positive charge in their structure), as they were carried to the MS by the residual EOF.

To overcome any influence from the different salt concentration of the purified samples on the amount injected, pressure injection was applied instead of electrokinetic injection being in all the cases the injected sample plug of 2.7 cm (3% of the capillary length). We have also taken into account that due to the complexity of the sample, during the CE-MS analysis it is possible that some metabolites can comigrate which disturbs the ionization conditions since the presence of strongly ionizable compounds will suppress the signal of the less ionizable ones. On the other hand, a disadvantage of CE-MS methods compared to LC-MS or GC-MS methods is the lower repeatability of peak areas. Nevertheless, good repeatability was obtained using this CE-MS method with %RSD values lower than 12% in the intra-day (n=5 injections) and inter-day (3 days, n=15 injections) repeatability study of ten different metabolites arbitrarily selected from the CE-MS electropherogram (peak area values and migration times are given as supporting information in Table S1).

3.3. CE-ESI-TOF MS metabolic analysis of human HT29 colon cancer cells

3.3.1.Matrix effect

After metabolite purification, the four different extracts from the HT29 colon cancer cells were analyzed by CE-MS. Although injected sample volume was the same in all the cases (3% of the capillary length), sample conductivity was expected to be different depending on the sample matrix due to their different salt and metabolite content. In order to study this effect on the CE-MS metabolite profile, the following experiments were carried out. First, formation of sodium formate clusters during CE-MS analysis of desalted (ABN extract) and non-desalted (MeOH extract) samples was observed

(extracted ion electropherograms, EIEs, are given in Figure S1 as supporting information). Basically, the results showed a small narrow peak in the desalted sample, while a broad band migrating from minutes 4 to 7 was observed in the non-desalted sample. It was also confirmed that the width of this band was proportional to the injected sample volume (data not shown). Typical sodium formate clusters observed were Na(NaCOOH)1 (90.9766 m/z), Na(NaCOOH)2 (158.9641 m/z), Na(NaCOOH)3 (226.9515 m/z), Na(NaCOOH)4 (294.9389 m/z), Na(NaCOOH)5 (362.9263 m/z), Na(NaCOOH)6 (430.9138 m/z), Na(NaCOOH)7 (498.9012 m/z), and Na(NaCOOH)8 (566.8886 m/z), which are produced by the interaction of the sodium ion from the sample with the formate counterions from the BGE. This clusters formation during the CE separation is also expected to have some influence on the CE-MS metabolite profile obtained for these samples. In order to study this effect, EIEs of common metabolites (identical m/z) to all extracts were represented (results are given as supporting information in Figure S2, indicating with letters from A to G the main obtained peaks). As expected, a migration time shift among the different extracts (ABN, C18, MeOH and ultrafiltration) was obtained. This effect is dependent on the ionic strength of the sample, which leads to different electrical conductivities and, as a result, different effective electric fields inside the capillary. In addition to the observed migration time shift, a slight increase in plate number of the peaks for the high-ionic-strength samples (non-desalted MeOH and ultrafiltration extracts) was obtained. This effect is contrary to the expected from theory (higher electromigration dispersion in the more salty samples) and although at the moment we do not see a clear reason, it could be explained through a stacking process induced by the sodium clusters band mentioned above [16].

3.3.2. Comparison of metabolite extraction methods

Base peak electropherograms (BPE) from the four different metabolite extracts were obtained by CE-MS (they are shown as supporting information in Figure S3). Comparing the two extracts obtained from SPE protocols (using ABN and C18 sorbents) it was observed that both CE-MS profiles were very similar. Namely, 80 compounds were detected by CE-MS in the ABN extract and 71 compounds in the C18 extract, observing that 62 of these compounds were common to both extracts. It is interesting to highlight at this point the better metabolite extraction power of the ABN sorbent at the selected SPE equilibration/elution conditions based on the higher number of metabolites detected (80) compared to the number obtained after C18 extraction (71).

Regarding the comparison of the metabolite extraction when only sample deproteinization was carried out (MeOH protein precipitation *vs*. ultrafiltration), the total number of compounds identified were 83 in the MeOH extract and 74 after ultrafiltration. Among all these compounds, 72 were common species. This result demonstrates the similarity of both deproteinization procedures. It is important to remark that sample treatment using protein precipitation with MeOH is simpler and less expensive protocol than ultrafiltration.

Next, selectivity of metabolite purification was studied comparing the best two protocols (which allowed the detection of a higher number of metabolites), namely, deproteinization and desalting using ABN cartridge, and MeOH deproteinization. After mass spectra analysis, important differences in terms of the selectivity of metabolite extraction were observed. In Figure 1, the EIEs of the compounds with different migration times observed in both extracts (ABN and methanolic extract) are

represented. Taking into account both extracts, a total number of 115 different metabolites were detected, of which 80 metabolites were found in ABN, 83 in MeOH extracts, being 48 compounds common to both extracts. In Figure 2, a bar plot of peak areas of the common metabolites extracted using the ABN cartridge or MeOH protein precipitation is presented. On the top of each column the metabolite (or peak) number assigned in Figure 1 and Table 1 is indicated. From this group of 48 compounds, statistically significant higher intensities (after ANOVA at 5% significance level) were observed for 22 peaks (namely, 1, 8, 22, 26, 33, 34, 50, 51, 53, 56, 65, 67, 72, 76, 77, 83, 86, 88, 90, 91, 96 and 97) when ABN extraction was used. On the other hand, among the common 48 compounds, only 8 showed a significantly higher ($p<0.05$) signal in the MeOH extract (peaks 17, 30, 52, 55, 57, 79, 84 and 110), what seems to indicate some clear bias in the quantitative extraction of these compounds depending on the purification protocol applied. Regarding the non-common compounds, 32 metabolites were detected using SPE with the ABN column, and 35 metabolites could only be detected by sample deproteinization with MeOH. This result is indicative of an important bias in the quantitative metabolomic analysis induced by the sample preparation step, as well as in the type and nature of the identified metabolites as will be corroborated below. The reproducibility obtained after metabolite purification using ABN or MeOH was similar. Namely, RSD values for peak areas were lower than 12% and 13% after ABN and MeOH extraction, respectively. These values were calculated for metabolites 56, 72 and 88 in three different extracts obtained under the same conditions using ABN or MeOH purification and analyzed in triplicate in the same day (total n=9). This information is given as supporting information in Table S2.

3.3.3. Metabolite identification

In Table 1, a list of detected compounds and tentative identification using both ABN and MeOH purification protocols is presented. The TOF-MS analyzer used in this work was operated to acquire spectra in the range of 50-600 m/z. Most of the observed ions were single protonated ions ($M+H^+$) and few of the species were detected as doubly charged (see Table 1). A total number of 115 different metabolites were detected with molecular masses from 90.0 to 777.2 Da. Among them, 45 compounds could not be identified since each one of the obtained m/z values was associated with more than 10 molecular formulas (considering an error lower than 10 ppm). In other cases a molecular formula was proposed but no endogenous metabolite could be found in metabolite databases. In certain cases, two or three metabolites were associated to a single molecular formula (metabolites 21, 29, 32, 38, 39, 52, 57, 64, 66, 70, 85, 108). In these cases, confirmation of identification was carried out based on their expected electrophoretic mobility at the separation pH and/or, when available, using standards. Thus, identification of histidine (peak 15), arginine (peak 20), lysine (peak 21), GABA (peak 29), valine (peak 52), serine (peak 53), leucine (or isoleucine) (peak 57), threonine (peak 64), glutamine (peak 66), proline (peak 70), tyrosine (peak 78) and aspartic acid (peak 85) was corroborated mixing the corresponding metabolite extract with standards solution. In certain cases, expected electrophoretic mobility was not enough information, and standard compound was not available in our laboratory. Thus, more than one metabolite had to be associated to a unique molecular formula. As an example, peak 32 with m/z value of 152,1072 and formula $C_9H_{13}NO$ was tentatively identified as N-methyltyramine and N-methylphenylethanolamine. Similar situation was founded for peaks 29, 38, 39 and 108 (see Table 1).

It is interesting to mention that among all amino acids detected (12), only histidine, valine, serine and leucine (or isoleucine) were observed in both metabolite extracts obtained after MeOH deproteinization or ABN purification. A more hydrophobic amino acid derivative as lipoillysine was only found in the ABN extract what seems to indicate some trend (although not definitive) to low polarity amino acids when ABN cartridges are used. The rest of amino acids (arginine, lysine, GABA, threonine, glutamine, proline, tyrosine and aspartic acid) were only detected when methanol was used for purification. Other identified endogenous metabolites with high number of polar groups in their structure, such as gluthathione disulfide (peak 92), iminodiacetate (peak 103), AMP nucleotide (peak 108), and phosphocreatine (peak 115), could only be detected in the methanolic extract. Finally, most metabolites with phosphorus, sulphur or at least four oxygen atoms presented the highest migration times in good agreement with their expected electrophoretic mobilities at the separation pH. Moreover, a good number of tripeptides (peaks 8, 12, 14, 60, 61, 65, 74, 88 and 91) were proposed. Due to the lack of information from MS/MS spectra in this work, only amino acidic composition is indicated for each peptide.

It was also observed the presence of three exogenous compounds, namely 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), bestatin and N-[N-(L-3-trans-carboxyirane-2-carbonyl)-L-leucyl]-agmatine (E-64) coming from the protease inhibitor cocktail used during cytosolic content isolation protocol. This finding is an additional demonstration on the consistency of these results.

In resume, from the 115 detected compounds, 44 metabolites were tentatively identified and classified in several groups. Most abundant compounds were amino acid and

polypeptides, although other compounds such as lipoic acid derivatives, steroid derivatives, carbohydrates derivatives, amino alcohol, amino acid phosphates, nucleosides, quaternary amines, pterins, acyl phosphates and nucleotides, were also found. Identified metabolites were automatically mapped to KEGG pathways using MassTrix. Among all tentatively identified compounds, 20 of them were related to the following metabolic pathways: (1) glycan biosynthesis and metabolism, (2) organismal systems: excretory system, (3) neuroactive ligand-receptor interaction, (4) carbohydrate metabolism, (5) lipid metabolism, (6) nucleotide metabolism, (7) energy metabolism, (8) metabolism of cofactors and vitamins, (9) environmental information processing, (10) genetic information processing and (11) amino acid metabolism. A particular case was creatine (peak 38), detected only in the methanolic extract, that was associated to more than 100 metabolic pathways. In Figure 3, a bar plot representing the number of metabolites associated to the above mentioned metabolic pathways, is presented. Information from both ABN and methanolic extracts is presented in order to show a general overview of the metabolic information obtained depending on the selected metabolite purification approach. As can be seen in Figure 3, metabolites from both ABN and methanolic extracts were associated to nine metabolic pathways (numbered from 3 to 11), although the number of metabolites was different in practically all cases. For instance, in the metabolic pathway number 11 of Figure 3, 16 metabolites were found in methanolic extract and only 5 in ABN extract. As a general trend in this study, the use of methanol deproteinization brought about a higher number of metabolites associated to known metabolic pathways and, therefore, wider metabolomic information could be obtained.

4. CONCLUDING REMARKS

Sample preparation is frequently underestimated in most metabolomic works. We have unequivocally demonstrated in this work that the composition and the quantity of metabolites detected depend to a large extent on the sample preparation step. Metabolite purification through methanol extraction showed good potential for metabolome characterization of human HT29 colon cancer cells, while the SPE results showed good extraction efficiency with different selectivity compared to protein methanol precipitation and ultrafiltration. However, SPE is usually considered more attractive for the on-line coupling extraction with CE-MS since sample manipulation can be simplified overall analytical procedure, allowing automatization of the sample treatment, separation and detection in metabolomic studies. Selection of an appropriate sample treatment for a certain metabolomic study is, therefore, crucial. These results show the important influence from the metabolite purification strategy since it can bias and in some cases mislead the conclusions achieved by Metabolomics. So, the question remains: is Metabolomics approachable?. Clearly, to achieve a whole Metabolomics study of a biological system is still challenging at least based on the current available methodologies. A possible solution could be the use of multiple sample preparation procedures to cover a broad (and, therefore, more informative and representative) range of metabolites and concentrations.

ACKNOWLEDGEMENTS

This work was supported by AGL2008-05108-C03-01 and 200870I185 (Ministerio de Ciencia e Innovación, Spain), and CSD2007-00063 FUN-C-FOOD (Programa CONSOLIDER, Ministerio de Educacion y Ciencia, Spain). C.I. thanks the Ministerio de Ciencia e Innovación for her FPI pre-doctoral fellowship.

REFERENCES

[1] Shulaev V. *Brief. Bioinform*. 2006, *7*, 128-139.

[2] Dettmer, K., Aronov, P.A., Hammock, B.D. *Mass Spectrom. Rev*. 2007, *26*, 51-78.

[3] Xiayan, L., Legido-Quigley, C. *Electrophoresis* 2008, *29*, 3724-3736.

[4] Garcia, D.E., Baidoo, E.E., Benke, P.I., Pingitore, F., Tang, Y.J., Villa, S., Keasling, J.D. *Curr. Opinion Microbiol*. 2008, *11*, 233-239.

[5] Issaq, H.J., Van, Q.N., Waybright, T.J., Muschik, G.M., Veenstra, T.D. *J. Sep. Sci*. 2009, *32*, 2183-2199.

[6] Monton, M.R.N., Soga, T. *J. Chromatogr. A* 2007, *1168*, 237-246.

[7] Harada, K., Fukusaki, E. *Plant Biotechnology* 2009, *26*, 47-52.

[8] Ramautar, R., Somsen, G.W., de Jong, G.J. *Electrophoresis* 2009, *30*, 276-291.

[9] Ramautar, R., Mayboroda, O.A., Somsen, G.W., de Jong, G.J. Electrophoresis 2011, 32, 52-65.

[10] Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J.A., Lim, E., Sobsey, C.A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H.J., Forsythe, I. *Nucleic Acids Res*. 2009, *37*, D603-D610.

[11] Smith, C.A., O'Maille, G., Want, E.J., Qin, C., Trauger, S.A., Brandon, T.R., Custodio, D.E., Abagyan, R., Siuzdak, G. Proceedings of the 9th International Congress of Therapeutic Drug Monitoring and Clinical Toxicology, Louisville, Kentucky. April 23-28, 2005.

[12] Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., Hirakawa, M. *Nucleic Acids Res*. 2010, *38*, D355-D360.

[13] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M. *Nucleic Acids Res.* 2006, *34*, D354-357.

[14] Kanehisa, M., Goto, S. *Nucleic Acids Res*. 2000, *28*, 27-30.

[15] Suhre, K., Schmitt-Kopplin, P. *Nucleic Acids Res.* 2008, *36*, W481-W484.

[16] Timerbaev, A.R., Hirokawa, T. Electrophoresis 2006, 27, 323-340.

FIGURE LEGENDS

Figure 1. CE-MS extracted ion electropherograms of the 115 m/z values detected in metabolite extracts obtained from SPE (ABN cartridge) or methanol deproteinization. CE-MS conditions are described in the text.

Figure 2. Representation of bar plot of peak areas of the common metabolites extracted using SPE (ABN cartridge) or protein precipitation with methanol. Metabolite number is assigned in Figure 1 and Table 1.

Figure 3. Number of compounds identified in some metabolic pathways in metabolite extracts obtained from SPE (ABN cartridge) or methanol deproteinization.

Table 1. Tentative identification of metabolites after CE-MS analysis.

No.	ABN	MeOH	m/z	z	Sigma™	Error (ppm)	Formula	Tentative identification	Classification	Database	KEGG pathway
1	YES	YES	145,0167	1	0,156	-2,9	$C_5H_5ClN_2O$				
2	YES	YES	90,9768	1	0,001	0,2	CNa_2O_2	Disodium formate	Inorganic salt		
3	YES	YES	272,9392	1	‹0.1	‹10	-[a)]				
4	YES	YES	189,0972	1	0,102	2,4	$C_6H_{12}N_4O_3$	X[b)]			
5	NO	YES	225,1431	1		+10					
6	YES	NO	427,2860	1	‹0.1	‹10	-				
7	YES	NO	134,9864	1	0,025	-4,1	$C_2H_2N_2O_3S$	X			
8	YES	YES	321,1257	1	0,026	-9,2	$C_{11}H_{20}N_4O_5S_1$	Tripeptide (G,N,M; A,Q,C)	Polypeptide	METLIN	
9	NO	YES	116,0837	1		+10					
10	YES	NO	335,1442	1	0,026	4,6	$C_{14}H_{26}N_2O_3S_2$	Lipoyllysine	Lipoic acid derivative	HMDB12996	
11	YES	NO	363,1763	1	0,04	3,6	$C_{16}H_{22}N_6O_4$	Thyrotropin-releasing factor	Polypeptide	HMDB05763	hsa04080
12	YES	NO	175,0855	2	0,020	4,6	$C_{12}H_{24}N_6O_4S_1$	Tripeptide (C,R,A)	Polypeptide	METLIN	
13	YES	NO	377,1924	1	0,024	9,1	$C_{21}H_{28}O_6$	18-Oxocortisol	Steroid derivative	HMDB00332	
14	YES	NO	203,1115	2	0,027	-6,2	$C_{20}H_{28}N_4O_5$	Tripeptide (I/L, W, S; W, T, V)	Polypeptide	METLIN	
15	YES	YES	156,0767	1	0,175	0,7	$C_6H_9N_3O_2$	Histidine	Amino acid	HMDB00177[c)]	hsa00340, hsa00410, hsa02010, hsa00970,
16	YES	NO	211,0943	1	‹0.1	‹10	-				
17	YES	YES	128,0818	1	0,006	-0,5	$C_5H_9N_3O$	X			
18	YES	NO	470,1381	1	‹0.1	‹10	-				
19	YES	NO	242,5777	2	‹0.1	‹10	-				
20	NO	YES	175,1193	1	0,046	-1,9	$C_6H_{14}N_4O_2$	Arginine	Amino acid	HMDB03416[c)]	hsa00472, hsa00330, hsa02010, hsa00970
21	NO	YES	147,1133	1	0,042	-3,5	$C_6H_{14}N_2O_2$	Lysine (+2 metabolites)	Amino acid	HMDB00182[c)]	hsa00780, hsa00310, hsa00300, hsa02010, hsa00970
22	YES	YES	504,1611	1	‹0.1	‹10	-				
23	YES	NO	546,2033	1	0,088	-0,9	$C_{20}H_{35}NO_{16}$	Lacto-N-triaose	Carbohydrate	HMDB06592	
24	YES	NO	194,0089	1	‹0.1	‹10	-				
25	YES	NO	532,1955	1	‹0.1	‹10	-				
26	YES	YES	160,1078	1	0,003	1,6	$C_6H_{13}N_3O_2$	X			
27	NO	YES	526,1438	1	‹0.1	‹10	-				
28	NO	YES	510,1496	1	‹0.1	‹10	-				
I.S.	YES	YES	138,0917	1	0,004	-2,9	$C_8H_{11}NO$	Tyramine			
29	NO	YES	104,0704	1	0,005	2,2	$C_4H_9NO_2$	Aminobutyric acid(+7 metabolites)	Amino acid	HMDB00112[c)]	
30	YES	YES	204,0484	1	0,004	2,4	$C_8H_{10}FNO_2S$	AEBSF (protease inhibitor)		CID 1701	
31	YES	NO	489,4360	1	‹0.1	‹10	-				
32	YES	YES	152,1072	1	0,058	-1,1	$C_9H_{13}NO$	N-Methyltyramine N-Methylphenylethanolamine	Amino alcohol Amino alcohol	HMDB03633 HMDB01387	hsa00350
33	YES	YES	218,0661	1	0,020	-0,6	$C_8H_{11}NO_6$	2-(Acetamidomethylidene)-3-(hydroxymethyl)butanedioc acid			
34	YES	YES	130,1592	1	0,007	-1,3	$C_8H_{19}N$	X			
35	YES	NO	228,1006	1	‹0.1	‹10	-				
36	YES	NO	230,0851	1	‹0.1	‹10	-				
37	YES	NO	244,0950	1	0,574	-9,0	$C_9H_{13}N_3O_5$	Cytidine	Nucleoside	HMDB00089	hsa00240
38	NO	YES	132,0771	1	0,007	-2,6	$C_4H_9N_3O_2$	Creatine Beta-Guanidinopropionic acid		HMDB00064 HMDB13222	hsa00260, hsa00330
39	NO	YES	264,0527	1	0,026	3,4	$C_9H_{13}NO_6S$	Epinephrine sulfate N-acetyl-S-(3-oxo-3-carboxy-n-propyl)cysteine		HMDB01876 HMDB02194	
40	YES	YES	303,6563	2	‹0.1	‹10	-				
41	YES	NO	188,2022	1	0,014	-1,8	$C_{11}H_{25}NO$	X			
42	YES	NO	174,1836	1	0,005	-5,6	$C_{10}H_{23}NO$	X			
43	NO	YES	218,1380	1	0,033	3,1	$C_{10}H_{19}NO_4$	Propionylcarnitine	Quaternary amine	HMDB00824	
44	YES	YES	192,1607	1	0,03	-1,4	$C_9H_{21}NO_3$	X			
45	YES	NO	271,1241	2	‹0.1	‹10	-				
46	YES	YES	548,1568	1	‹0.1	‹10	-				
47	YES	YES	200,2367	1	0,007	2,7	$C_{13}H_{29}N$	Octamylamine		CID 10406	
48	YES	NO	214,2522	1	0,008	3,5	$C_{14}H_{31}N$	X			
49	YES	YES	218,2083	1	0,133	2,2	$C_8H_{23}N_7$	X			
50	YES	YES	287,1039	1	0,010	6,4	$C_8H_{14}N_8O_2S$	X			
51	YES	YES	230,2475	1	0,02	1,7	$C_{14}H_{31}NO$	X			
52	YES	YES	118,0868	1	0,005	-4,4	$C_5H_{11}NO_2$	Valine (+ 3 metabolites)	Amino acid	HMDB00883[c)]	hsa00640, hsa00770, hsa00280, hsa00290, hsa00970
53	YES	YES	106,0499	1	0,021	0,2	$C_3H_7NO_3$	Serine	Amino acid	HMDB00187[c)]	hsa00600, hsa00920, hsa00460, hsa00270, hsa00260, hsa00970
54	NO	YES	301,1172	1	‹0.1	‹10	-				
55	YES	YES	387,0834	1	0,019	2,3	$C_{12}H_{10}N_{12}O_2S$	X			
56	YES	YES	349,1170	1	0,018	-7,8	$C_{15}H_{16}N_4O_6$	Riboflavin reduced	Pterin	HMDB01557	
57	YES	YES	132,1015	1	0,004	2,8	$C_6H_{13}NO_2$	Leucine or Ilsoleucine (+ 3 metabolites)	Amino acid	HMDB00687, HMDB00172[c)]	hsa00280, hsa00290, hsa00970
58	NO	YES	299,1036	1	‹0.1	‹10	-				
59	YES	NO	335,1377	1	0,024	1,9	$C_{12}H_{22}N_4O_5S$	Tripeptide (V, N, C; M, Q, G; A, M, N)	Polypeptide	METLIN	
60	YES	NO	429,1509	1	‹0.1	‹10	-				

61	YES	NO	363,1681	1	0,023	4,5	$C_{14}H_{26}N_4O_5S_1$	Tripeptide (I/L,C,Q; N,V,M)	Polypeptide	METLIN	
62	YES	NO	258,2788	1	0,011	1,3	$C_{16}H_{35}NO$	X			
63	YES	NO	288,2892	1	0,037	1,9	$C_{17}H_{37}NO_2$	X			
64	NO	YES	120,0659	1	0,737	-3,5	$C_4H_9NO_3$	Threonine (+ 2 metabolites)	Amino acid	HMDB00167 [c]	hsa00860, hsa00290, hsa00260, hsa00970
65	YES	YES	291,1682	1	0,030	-6,4	$C_{11}H_{22}N_4O_5$	Tripeptide (G,S,K)	Polypeptide	METLIN	
66	NO	YES	147,0758	1	0,006	4,1	$C_5H_{10}N_2O_3$	Glutamine (+ 2 metabolites)	Amino acid	HMDB00641 [c]	hsa00240, hsa04964, hsa00230, hsa00471, hsa00910, hsa00250, hsa00330, hsa02010, hsa00970
67	YES	YES	323,1961	1	0,018	1,3	$C_{17}H_{26}N_2O_4$	X			
68	YES	NO	305,1858	1	0,012	5,9	$C_{15}H_{29}O_4P$	Dolichol phosphate	Acyl phosphate	HMDB06353	hsa00510
69	YES	NO	351,2254	1	‹0.1	‹10	-				
70	NO	YES	116,0714	1	0,0029	-6,9	$C_5H_9NO_2$	Proline (+ 1 metabolite)	Amino acid	HMDB00162 [c]	hsa00330, hsa02010, hsa00970
71	YES	YES	309,1806	1	0,009	0,9	$C_{16}H_{24}N_2O_4$	Bestatin (protease inhibitor)		CID 72172	
72	YES	YES	323,1964	1	0,110	0,6	$C_{17}H_{26}N_2O_4$	X			
73	NO	YES	365,0834	1	‹0.1	‹10	-				
74	YES	YES	327,1417	1	0,1885	-2,2	$C_{12}H_{18}N_6O_5$	Tripeptide (N,H,G)	Polypeptide	METLIN	
75	YES	NO	408,1959		‹0.1	‹10	-				
76	YES	YES	380,6276	2	‹0.1	‹10	-				
77	YES	YES	474,1998	1	‹0.1	‹10	-				
78	NO	YES	182,0819	1	0,035	-4,2	$C_9H_{11}NO_3$	Tyrosine	Amino acid	HMDB00158 [c]	hsa00130, hsa00360, hsa00400, hsa00730, hsa00410, hsa00970
79	YES	YES	570,1186	1	‹0.1	‹10	-				
80	YES	YES	397,2848	1	‹0.1	‹10	-				
81	YES	NO	584,1268	1	‹0.1	‹10	-				
82	NO	YES	191,1025	1	0,005	0,6	$C_7H_{14}N_2O_4$	Diaminopimelic acid	Amino acid	HMDB01370	hsa00300
83	YES	YES	474,1990	1	‹0.1	‹10	-				
84	YES	YES	413,2810	1	‹0.1	‹10	-				
85	NO	YES	134,0451	1	0,0055	-2,5	$C_4H_7NO_4$	Aspartic acid (+ 1 metabolite)	Amino acid	HMDB00191 [c]	hsa00760, hsa04080, hsa00340, hsa00410, hsa00460, hsa00770, hsa00910, hsa00250, hsa00270, hsa00300, hsa00260, hsa00330, hsa02010, hsa00970
86	YES	YES	429,3182	1	0,004	0,3	$C_{20}H_{40}N_6O_4$	X			
87	YES	YES	445,3136	1	‹0.1	‹10	-				
88	YES	YES	492,2153	1	0,028	-4,9	$C_{27}H_{29}N_3O_6$	Tripeptide (Y,Y,F)	Polypeptide	METLIN	
89	NO	YES	510,1815	1	0,159	2,1	$C_{21}H_{36}NO_9S_2$	X			
90	YES	YES	443,2985	1	‹0.1	‹10	-				
I.S.	NO	YES	182.0478	1	0,005	1,9	$C_5H_{11}NO_4S$	Methionine sulfone			
91	YES	YES	394,1854	1	0,106	-6,3	$C_{16}H_{31}N_3O_4S_2$	Tripeptide (I/L,M,M)	Polypeptide	METLIN	
92	NO	YES	307,0835	2	0,016	-0,8	$C_{20}H_{32}N_6O_{12}S_2$	Glutathione disulfide	Polypeptide	HMDB03337	hsa0480
93	NO	YES	400,1326	1	0,003	-3,8	$C_{11}H_{21}N_5O_{11}$	X			
94	NO	YES	396,1639	1	‹0.1	‹10	-				
95	NO	YES	290,1019	2	‹0.1	‹10	-				
96	YES	YES	298,1022	2	‹0.1	‹10	-				
97	YES	YES	389,6241	2	‹0.1	‹10	-				
98	YES	YES	381,6264	2	‹0.1	‹10	-				
99	NO	YES	326,6199	2	‹0.1	‹10	-				
100	YES	YES	358,2086	1	0,024	-0,3	$C_{15}H_{27}N_5O_5$	E-64 (protease inhibitor)		CID 123985	
101	NO	YES	258,2770	1		+10					
102	NO	YES	326,6191	1		+10					
103	NO	YES	134,0455	1	0,017	-5,6	$C_4H_7NO_4$	Iminodiacetate			
104	NO	YES	217,0822	1	‹0.1	‹10	-				
105	NO	YES	186,0584	1	0,005	-0,60	$C_8H_{11}NO_2S$	X			
106	YES	YES	369,0941	1	‹0.1	‹10	-				
107	YES	YES	552,1293	1	‹0.1	‹10	-				
108	NO	YES	348,0731	1	0,012	-7,8	$C_{10}H_{14}N_5O_7P$	2'-Deoxyguanosine 5'-monophosphate Adenosine 2'-phosphate Adenosine 5'-phosphate	Nucleotide	HMDB01044 HMDB11617 HMDB00045	hsa01100 [d]
109	NO	YES	184,0743	1	0,003	-9,6	$C_5H_{14}NO_4P$	Phosphocholine		METLIN	
110	YES	YES	478,1304	1	‹0.1	‹10	-				
111	YES	NO	516,0779	1	‹0.1	‹10	-				
112	YES	NO	529,3531	1	‹0.1	‹10	-				
113	NO	YES	293,0966	1	‹0.1	‹10	-				
114	NO	YES	504,1329	1	‹0.1	‹10	-				
115	NO	YES	212,0426	1	0,034	2,1	$C_4H_{10}N_3O_5P$	Phosphocreatine	Amino acid phosphate	HMDB01511	hsa01100 [d]
I.S.	YES	YES	303.0607	1	0,006	2,7	$C_8H_{18}N_2O_6S_2$	PIPES			

a) More than 10 molecular formulas were assigned to the obtained m/z value.
b) No endogenous metabolite can be associated to the assigned molecular formula
c) Tentative identification was confirmed with commercial standards
d) hsa01100 code includes more than 20 metabolic pathways

Supporting information

Figure S1. CE-MS extracted ion electropherograms of the sodium formate clusters obtained from ABN extract (upper electropherogram) and methanolic extract (lower electropherogram). Characteristic mass spectra are shown in the lower panel. CE-MS conditions are described in "CE-ESI-TOF MS conditions" section.

Figure S2. CE-MS extracted ion electropherograms of the 43 m/z ions common to all extracts. Seven peaks (from A to G) are marked in the electropherograms for a better comparison of the purification strategies. Peak A: 122.08 m/z, peak B: 204.05 m/z, peak C: 387.08 m/z, peak D: 291.17 m/z, peak E: 309.18 m/z, F: 413.29 m/z, and G: 182.05 m/z. CE-MS conditions are described in "CE-ESI-TOF MS conditions" section.

Figure S3. Base peak electropherograms of the cytosolic fraction from colon cancer cell culture after four different metabolite extraction procedures. CE-MS conditions are described in "CE-ESI-TOF MS conditions" section.

Table S1. Peak areas, migration times and RSD values for ten different metabolites detected by CE-MS after ABN extraction in three different days. CE-MS conditions are described in section 2.5.

DAY 1	R1		R2		R3		R4		R5		t_{mig}			Area		
Metabolite No.	t mig	A	t mig	A	t mig	A	t mig	A	t mig	A	Med	SD	RSD	Med	SD	RSD
15	6.55	8265	6.62	7654	6.75	8038	6.85	7652	6.97	8017	6.75	0.17	2.5	7925	267	3.4
22	7.05	78569	7.15	72086	7.29	74241	7.37	75540	7.50	82332	7.27	0.18	2.4	76554	3993	5.2
32	7.57	725663	7.69	702632	7.86	767788	7.95	715422	8.09	763530	7.83	0.21	2.6	735007	29186	4.0
40	8.16	14532	8.22	14628	8.45	15550	8.55	15201	8.71	14123	8.42	0.23	2.7	14807	566	3.8
46	8.70	4205	8.77	3955	8.93	4475	9.01	4503	9.19	4355	8.92	0.20	2.2	4299	225	5.2
56	9.40	110523	9.46	116568	9.65	112865	9.76	102568	9.94	109936	9.64	0.22	2.3	110492	5139	4.7
72	10.35	2685020	10.41	2638364	10.58	2890925	10.71	2769980	10.88	2432013	10.59	0.22	2.1	2683260	170139	6.3
88	11.15	130982	11.23	127066	11.39	141619	11.52	146230	11.68	125969	11.39	0.21	1.9	134373	9064	6.7
91	11.51	2664120	11.70	2874057	11.79	2859792	11.89	2669031	12.16	2508991	11.81	0.24	2.0	2715198	152811	5.6
106	16.11	335642	17.53	308625	17.40	381204	16.90	370219	17.91	359821	17.17	0.70	4.0	351102	29116	8.3
DAY 2	**R1**		**R2**		**R3**		**R4**		**R5**		**t_{mig}**			**Area**		
Metabolite No.	**t mig**	**A**	**t mig**	**A**	**t mig**	**A**	**t mig**	**A**	**t mig**	**A**	**Med**	**SD**	**RSD**	**Med**	**SD**	**RSD**
15	6.12	8803	6.22	9288	6.32	8166	6.44	8456	6.69	9113	6.36	0.22	3.5	8765	461	5.3
22	6.65	65684	6.63	68612	6.79	59917	6.93	62464	7.21	64121	6.84	0.24	3.5	64160	3280	5.1
32	7.24	700996	7.17	741936	7.35	643436	7.55	675458	7.85	693140	7.43	0.27	3.7	690993	36065	5.2
40	7.79	17698	7.62	16398	7.86	14854	8.08	16584	8.40	16918	7.95	0.30	3.8	16490	1041	6.3
46	8.28	4809	8.00	5034	8.29	4350	8.54	4583	8.87	4664	8.40	0.33	3.9	4688	255	5.4
56	8.90	95684	8.52	98724	8.85	85245	9.14	89878	9.49	91184	8.98	0.36	4.0	92143	5232	5.7
72	9.75	2515102	9.26	2587314	9.65	2220188	9.97	2355491	10.33	2384463	9.79	0.40	4.0	2412512	143272	5.9
88	10.42	158847	9.82	167502	10.28	143237	10.63	152494	11.00	153825	10.43	0.44	4.2	155181	8902	5.7
91	10.86	2375011	10.19	2488590	10.70	2112414	11.05	2265612	11.43	2278126	10.85	0.46	4.2	2303951	139610	6.1
106	14.82	434226	13.92	452856	14.65	379092	15.10	412280	15.58	410724	14.81	0.61	4.1	417836	27738	6.6
DAY 3	**R1**		**R2**		**R3**		**R4**		**R5**		**t_{mig}**			**Area**		
Metabolite No.	**t mig**	**A**	**t mig**	**A**	**t mig**	**A**	**t mig**	**A**	**t mig**	**A**	**Med**	**SD**	**RSD**	**Med**	**SD**	**RSD**
15	5.85	9566	5.92	10005	6.02	9259	6.08	10082	6.25	9577	6.02	0.15	2.6	9698	342	3.5
22	6.33	63332	6.35	59379	6.43	56421	6.50	63777	6.67	59206	6.46	0.14	2.1	60423	3094	5.1
32	6.82	829796	6.87	880844	6.98	781408	7.00	846435	7.24	808650	6.98	0.16	2.3	829427	37652	4.5
40	7.24	16104	7.34	17867	7.42	17621	7.44	18864	7.69	17972	7.43	0.17	2.3	17686	1001	5.7
46	7.58	4991	7.74	4961	7.77	4492	7.80	5002	8.07	4853	7.79	0.18	2.3	4860	214	4.4
56	8.09	86017	8.31	85230	8.32	94152	8.32	82006	8.62	86626	8.33	0.19	2.3	86806	4476	5.2
72	8.72	2179522	9.07	2305314	8.99	2083717	8.96	2208366	9.35	2097898	9.02	0.23	2.5	2174963	89990	4.1
88	9.27	155015	9.70	158715	9.57	175790	9.52	177548	9.94	166363	9.60	0.25	2.6	166686	10009	6.0
91	9.59	2165933	10.05	2021198	9.89	2326637	9.84	2040365	10.30	2149446	9.93	0.26	2.6	2140716	122100	5.7
106	13.67	402900	14.48	437380	14.17	481275	13.92	454391	14.71	453070	14.19	0.42	2.9	445803	28707	6.4

Table S2. Overall reproducibility (sample preparation + CE-MS analysis) given as peak areas and RSD values for metabolites 56, 72 and 88 determined in three ABN and MeOH extracts obtained under the same conditions and analyzed in triplicate by CE-MS (total n=9).

ABN								
No.	**m/z**	**CE-MS Rep.**	**Peak area (Treatment 1)**	**Peak area (Treatment 2)**	**Peak area (Treatment 3)**	**Mean**	**SD**	**% RSD**
56	349.117	R1	105065	101583	108973	104504	6462	6
		R2	107374	92001	113464			
		R3	104004	98574	109500			
72	323.1964	R1	2296328	2359709	2465852	2395546	154976	6
		R2	2424419	2135371	2608870			
		R3	2389613	2269479	2610277			
88	492.2153	R1	130842	154736	162406	156981	18451	12
		R2	141178	154733	175277			
		R3	139635	164498	189523			
MeOH								
No.	**m/z**	**CE-MS Rep.**	**Peak area (Treatment 1)**	**Peak area (Treatment 2)**	**Peak area (Treatment 3)**	**Mean**	**SD**	**% RSD**
56	349.117	R1	23018	26045	24145	24658	1852	8
		R2	21283	27317	24878			
		R3	23553	25745	25939			
72	323.1964	R1	10958	10716	12283	10949	1129	10
		R2	9939	10175	11909			
		R3	9862	9879	12819			
88	492.2153	R1	130118	102388	95395	110122	14524	13
		R2	118505	96908	102439			
		R3	135152	101220	108975			