Research Article

Use of brightness wavelet transformation for automated analysis of serum metallothioneins- and zinc-containing proteins by Western blots to subclassify childhood solid tumours

In this study, we determined serum levels of metallothioneins (MTs) and zinc in children with solid tumours (neuroblastoma, Hodgkin lymphoma, medulloblastoma, osteosarcoma, Ewing sarcoma and nephroblastoma) by differential pulse voltammetry Brdicka reaction and ELISA. Zn(II) level in patients sera was 40% compared to controls, contrariwise, MT level was 4.2 × higher in patients. No significant differences among single diagnoses were found both for Zn(II) and MT. When determined Zn(II)/MT ratio, in controls its value was 24.6, but it was 2.6 in patients. After Western-blotting with anti-MT and anti-Zn chicken antibodies, variable intensities of the bands within the samples were observed. The brightness curve obtained for each sample both for MT- and Zn blots was further analysed to produce a list of band positions together with some complementary information related to the intensity of the observed bands by the optimised algorithm. We constructed from those profiles decision trees that enable to distinguish different groups of tumours. The blood samples were heat-treated, in which we supposed mainly MT, but samples contained other thermostable Zn-containing proteins that were helpful for identification of embryonal tumours with 88% accuracy and for identification of sarcomas with 78% accuracy. In MT blots the accuracies were 53 and 45%, respectively. Simultaneous analysis of MT and Zn blots did not increased accuracy of identification neither in embryonal tumours (80%) nor in sarcomas. Those results are promising not only from diagnostic point of view but particularly in the area of studying of individual MT isoforms and their aggregates in malignant tumours.

Keywords:
Cancer / Metallothionein / Signal processing / Zinc

DOI 10.1002/elps.201200561

1 Introduction

Metallothioneins (MTs) are evolutionary highly conserved ubiquitous metal-binding proteins, which were discovered by Margoshes and Valee as a cadmium-binding proteins in horse kidney in 1957 [1]. Later on, their involving in heavy metal homeostasis, oxidative stress coping, gene expression and transcription regulation, enzymes activation, apoptosis and cell proliferation have been found [2, 3]. Till now, four major isoforms (MT-1 through MT-4) have been identified in mammals [4, 5]. MTs genes are tightly linked, and at a minimum they consist of eleven MT-1 genes (MT-1A, -B, -E, -F, -G, -H, -I, -J, -K, -L and -X) encoding functional or non-functional RNAs, and one gene for each of the other MTs isoforms (the MT-2 A gene, MT-3 gene and MT-4 gene) [6]. Concerning their primary structure, they are rich in cysteine and have no aromatic amino acids [7]. Current knowledge of MTs is juxtaposed with our understanding of the pathogenesis of disease [8]. MT is known to modulate three fundamental processes: (i) the release of gaseous mediators such as hydroxyl radical or nitric oxide; (ii) apoptosis; (iii) the binding and exchange of metals such as zinc and copper on one side, and cadmium and platinum on the other side. Associations among MTs and several diseases, including cancer, circulatory and septic shock, coronary artery disease and Alzheimer’s disease have been found. Furthermore, strong evidence exists that MTs modulate the immune system [5].

Correspondence: Dr. Rene Kizek, Department of Chemistry and Biochemistry, Mendel University in Brno, Zemedelska 1, CZ-613 00 Brno, Czech Republic
E-mail: kizek@sci.muni.cz
Fax: +420-5-4521-2044

Abbreviation: MT, metallothionein

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

www.electrophoresis-journal.com
Great attention is paid to the study of MTs in tumours [9–12]. Most studies are focused on the immunohistochemical determination of MTs directly in tumours [12–21], but also on its determination in serum by various methods such as Brdicka reaction or ELISA [12, 19]. MT can serve as a prognostic marker in central nervous system tumours of childhood and adolescence [22, 23], osteosarcoma [24], breast cancer [25–27], pancreatic islet cells tumours [28] and tongue squamous cell carcinoma [29]. MT can also serve as a serum tumour marker in prostate cancer [30, 31], head and neck tumours [16], childhood solid tumours [19] and melanoma [18]. Expression of MT may help to distinguish between benign and malignant tumour, as shown at thyroid tumours [32], prostatic lesions [33], gastrointestinal stromal tumours and gastric carcinomas [34]. At isoform levels, MT expression in breast and prostate cancers, renal tumours and papillary thyroid cancer was reviewed by Thimoorthy et al., who found that expression of MT isoforms was down/upregulated differentially according to cancer type [35, 36].

Based on the abovementioned fact one may suggest that there must a mechanism differently influencing somewhat structure or concentration or both of MTs. In the light of this, it was found that MTs form aggregates, especially under higher MTs concentration, which can be observed as a change of their electrophoretic mobility [30, 37, 38]. In serum of patients with oncological disease, five times enhanced MT level [16, 19, 39] and decreased level of antioxidants and increased GSH/GSSG ratio (marker of oxidative stress) compared to controls were found [30, 40, 41]. Based on these results, it can be hypothesised that MTs occur in oligomeric form in cancer patients, and, moreover, the state is dependent on heavy metals content and antioxidant status of a patient [42]. Western-blotting is beneficial to be used for analysis of MTs aggregates in a complex sample like blood serum, unless other methods, which have been already used, for example capillary electrophoresis, gel electrophoresis, mass spectrometry and others [43–45]. The blots can be used for samples typing, which has been shown as Western-blot fingerprinting (“westprinting”) by cluster analysis UPGMA (Unweighted Pair Group Method with Arithmetic Mean) for characterisation of Pseudomonas [46] or Campylobacter strains [47] or identification of pathogenic protozoans Eimeria in birds [48].

The aim of this study was to automatically analyse Western-blot profiles of MT and Zn-proteins in serum of patients with childhood solid tumours and to determine the difference among single samples and diagnoses by advanced mathematical approaches.

2 Materials and methods

2.1 Chemicals

Rabbit liver MT (MW 7143 g/mol), containing 5.9% Cd and 0.5% Zn, was purchased from Sigma Aldrich (St. Louis, USA). Other chemicals used were also purchased from Sigma Aldrich unless stated otherwise. The stock standard solutions of MT (1 mg/mL) was prepared with ACS (American Chemical Society) grade water and stored in the dark at −20°C. Working standard solutions were prepared daily by dilution of the stock solutions with ACS water. All other solutions used were prepared in MilliQ water. Deionised water underwent de-mineralisation by reverse osmosis using the instrument Aqua Osmotic 02 (Aqua Osmotic, Tisnov, Czech Republic) and then subsequently purified using Millipore RG (Millipore, USA, 18 MΩ – MilliQ water. The pH value and conductivity was measured using inoLab Level 3 (Wissenschaftlich-Technische Werkstatten, Weilheim, Germany).

2.2 Preparation of anti-Zn and anti-MT antibodies

Chicken anti-zinc antibodies were prepared by HENA, Prague, Czech Republic. Two hens were immunised by Zn-keyhole limpet hemocyanin complex according to He et al. [49]. From the egg yolk the IgY fraction with reactivity to Zn-keyhole limpet hemocyanin was obtained. The antibodies were stabilised with 0.1% sodium azide in PBS. The protein concentration was 39.6 mg/mL in immunoglobulin fraction. Immunoreactivity and specificity of the antibodies was characterised by the manufacturer and by Krizkova et al. according to the published paper [20].

Chicken anti-MT antibodies were prepared by HENA. Two hens were immunised by the commercially available MT (1 mg of the mixture of horse MT 1 and MT 2, Sigma Aldrich), which was diluted in water and incubated for seven days at room temperature in order to polymerize. From the egg yolk, the IgY fraction with reactivity to MT was obtained. The antibodies in PBS were stabilised with 0.1% sodium azide. The protein concentration was 54.7 mg/mL in immunoglobulin fraction. Immunoreactivity and specificity of the antibodies was characterised by the manufacturer and by Krizkova et al. according to the published paper [42].

2.3 Human blood serum

Blood samples were obtained from 38 children hospitalised at Department of Paediatric Haematology and Oncology of Faculty Hospital Motol with newly diagnosed solid tumours (medulloblastoma (n = 10), neuroblastoma (n = 12), neuroblastoma HR (n = 6), osteosarcoma (n = 12), Ewing sarcoma (n = 6), Hodgkin lymphoma (n = 5), nephroblastoma (n = 5); average age 7.3 years). The blood samples were collected before chemo- and radiotherapy. The samples were primarily intended for routine biochemical tests at Department of Clinical Biochemistry and Pathobiology of Faculty Hospital Motol. Serum was separated by centrifugation at 4000 × g for 10 min. For further investigations sera unemployed for routine biochemical tests were used. The samples were stored in −80°C until assayed. Samples (blood) taking and subsequent processing was approved by Ethic Committee of Faculty Hospital Motol.
Concerning control samples, blood samples were obtained from six healthy volunteers (three males, three females; average age 32.5 years). Plasma was separated by centrifugation at 4000 × g for 10 min in plasma preparation tubes with heparin (Dialab, Czech Republic). The samples were stored in −80°C until assayed.

The samples were kept at 99°C in a thermomixer (Eppendorf 5430, Germany) for 15 min with shaking in order to remove ballast proteins and peptides, which could influence the electrochemical response [50]. The denatured homogenates were centrifuged at 4°C, 15 000 × g for 30 min (Eppendorf 5402).

2.4 Zn(II) and total proteins determination

Total proteins content in the samples was determined according to Bradford [51]. Briefly, 10 μL of sample diluted with 0.1 M phosphate buffer pH 7.6 was mixed with 190 μL of the Bradford reagent (0.117 mM CBB G 250, 50 mL of 96% ethanol, 100 mL of 85% phosphoric acid filled up with deionised water to total volume of 1 L) in a microplate (Immuno 96 MicroWell™ Solid Plates). Absorbance was measured at 595 nm, 10 μL of the phosphate buffer mixed with 190 μL of Bradford reagent was used as a blank. After 10 min incubation at room temperature the absorbance was read at 595 nm using a microplate reader (Infinite, TECAN, Japan).

For preparation of calibration curve, bovine serum albumin was used within the concentration range from 10 to 500 μg/mL. The equation was \( y = 1.037x - 0.009 \), \( R^2 = 0.9901 \).

2.5 Differential pulse voltammetry for Zn(II) determination

Determination of zinc by differential pulse voltammetry were performed with 797 VA Stand instrument connected to 813 Autosampler (Metrohm, Switzerland), using a standard cell with three electrodes. Method was adopted from [52, 53]. Briefly, a hanging mercury drop electrode with a drop area of 0.4 mm² was the working electrode. An Ag/AgCl/3M KCl electrode was the reference and platinum electrode was auxiliary. For data processing VA Database 2.2 by Metrohm, using a standard cell with three electrodes and cooled sample holder (4°C) according to previously published papers [16, 38, 50, 54]. Briefly, there was used hanging mercury drop electrode as the working electrode. An Ag/AgCl/3M KCl electrode was the reference and glassy carbon electrode was auxiliary. Brdicka supporting electrolyte containing 1 mM Co(NH₃)₆Cl₃ and 1 M ammonia buffer (NH₄Cl, pH 9.6) was used. The parameters of the measurement were as follows: initial potential of −0.7 V, end potential of −1.75 V, modulation time 0.057 s, time interval 0.2 s, step potential 2 mV and modulation amplitude 250 mV. All experiments were carried out at 4°C employing thermostat Julabo F25 (Laborteknik, Germany). The obtained calibration curve for MT was as follows \( y = 2.1713x - 1.9824; R^2 = 0.996, n = 5, R.S.D. = 2.4\% \). For data processing GPES 4.9 supplied by EcoChemie was employed.

2.6 Differential pulse voltammetry Brdicka reaction for determination of MTs

Differential pulse voltammetric measurements were performed with 747 VA Stand instrument connected to 746 VA Trace Analyser and 695 Autosampler (Metrohm), using a standard cell with three electrodes and cooled sample holder (4°C) according to previously published papers [16, 38, 50, 54]. Briefly, there was used hanging mercury drop electrode as the working electrode. An Ag/AgCl/3M KCl electrode was the reference and glassy carbon electrode was auxiliary. Brdicka supporting electrolyte containing 1 mM Co(NH₃)₆Cl₃ and 1 M ammonia buffer (NH₄Cl, pH 9.6) was used. The parameters of the measurement were as follows: initial potential of −0.7 V, end potential of −1.75 V, modulation time 0.057 s, time interval 0.2 s, step potential 2 mV and modulation amplitude 250 mV. All experiments were carried out at 4°C employing thermostat Julabo F25 (Laborteknik, Germany). The obtained calibration curve for MT was as follows \( y = 2.1713x - 1.9824; R^2 = 0.996, n = 5, R.S.D. = 2.4\% \). For data processing GPES 4.9 supplied by EcoChemie was employed.

2.7 SDS-PAGE and Western-blotting for proteins fingerprinting

SDS-PAGE was performed according to Laemmli [1] using a Mini Protean Tetra apparatus with gel dimension of 8.3 cm × 7.3 cm (Bio-Rad, USA). First 15% w/v running, then 5% w/v stacking gel was poured. The gels were prepared from 30% w/v acrylamide stock solution with 1% w/v bisacrylamide. The polymerization of the running or stacking gels was carried out at room temperature for 45 min or 30 min, respectively. Prior to analysis the samples were mixed with reduction (7.5% β-mercaptoethanol) sample buffer in a 2:1 ratio. The samples were boiled for 2 min, and then 4 μL of the sample was loaded onto a gel. For determination of the molecular mass, the protein ladder “Precision plus protein standards” from Biorad was used. The electrophoresis were run at 150 V for 1 h (Power Basic, Biorad USA) in Tris-glycine buffer (0.025 M Trizma-base, 0.19 M glycine and 0.0035 M SDS, pH 8.3) at room temperature (22°C). Silver staining of the gels was performed according to Oakley et al. [2].

Western-blotting: after the electrophoretic separation the proteins were transferred on a PVDF (Biorad, USA) membrane by using of Biometra Fastblot apparatus (Biometra, Germany). PVDF membranes were activated by soaking in 7.3% acrylamide solution with 1% w/v bisacrylamide. The polymerization of the running or stacking gels was carried out at room temperature for 45 min or 30 min, respectively. Prior to analysis the samples were mixed with reduction (7.5% β-mercaptoethanol) sample buffer in a 2:1 ratio. The samples were boiled for 2 min, and then 4 μL of the sample was loaded onto a gel. For determination of the molecular mass, the protein ladder “Precision plus protein standards” from Biorad was used. The electrophoresis were run at 150 V for 1 h (Power Basic, Biorad USA) in Tris-glycine buffer (0.025 M Trizma-base, 0.19 M glycine and 0.0035 M SDS, pH 8.3) at room temperature (22°C). Silver staining of the gels was performed according to Oakley et al. [2].
NaH$_2$PO$_4$ and 4.3 mM Na$_2$HPO$_4$, pH 7.4) for 30 min. The incubation either with anti-MT or anti-Zn chicken primary antibody in dilution of 1:500 in PBS with 0.1% of BSA was carried out for 12 h at 20°C. Specificities and cross-reactivity of the used antibodies were checked by blotting with IgY yolk fractions from hen’s eggs before immunization under the same conditions. After the three times repeated washing with PBS containing 0.05% v/v Tween-20 (PBS-T) for 5 min the membrane was incubated in the presence of secondary antibody (rabbit anti-chicken labelled with horseradish peroxidase, Sigma-Aldrich, in dilution 1:6000) for 1 h at room temperature. Then, the membrane was washed three times with PBS-T for 5 min and incubated with chromogenic substrate (0.4 mg/mL AEC (3-aminoethyl-9-carbazole) in 0.5 M acetate buffer with 0.1% H$_2$O$_2$, pH 5.5), after the adequate development the reaction was stopped by rinsing with water.

### 3 Results and discussion

#### 3.1 Determination of Zn and MT

MT and Zn$^{2+}$ content were determined by voltammetry and ELISA (Fig. 1). It was found that average Zn$^{2+}$ level was 6.0 ± 0.8 μM in patients’ sera and 15 ± 2 μM in controls. Average MT content in patients was 2.3 ± 0.2 μM determined by differential pulse voltammetry Brdicka reaction and 2.2 ± 0.1 μM determined by ELISA. In controls, the average MT content was 0.6 ± 0.4 μM determined by differential pulse voltammetry Brdicka reaction and 0.5 ± 0.3 μM determined by ELISA. MT and Zn$^{2+}$ levels for individual diagnosis are shown in Fig. 1. It clearly follows from the results that while Zn$^{2+}$ level in patients sera was approximately 40% compared to controls, contrariwise, MT was 4.2 × higher in patients compared to controls. No significant differences among single diagnoses were found both for Zn$^{2+}$ and MT. It is obvious that Zn$^{2+}$ content in patients with solid childhood tumours is decreased [55], which is in well agreement with the previously published studies, where zinc status has been studied in Hodgkin’s lymphoma, leukaemia, bone and brain tumours [40, 41, 56–59]. When determined Zn$^{2+}$/MT ratio, its value was 24.6 in controls, but the ratio was 2.6 in patients. This phenomenon can be associated with the degree of MTs polymerization, i.e. change of zinc binding capacity and therefore change of their functions in patients’ sera can be questioned [60].

Based on the possible polymerisation of MTs, we investigated this phenomenon in this study. After blotting with chicken anti-MT antibodies, six bands with approximate molecular weights 10, 17, 19, 25, 35 and 50 kDa were observed. The intensity and the presence of the bands varied within the group. In addition to these bands, additional bands in variable molecular weights were observed according to sample origin. These bands correspond to MTs isoforms and their high-molecular aggregates and their intensity corresponds to MT concentration in the sample [42].

After blotting of samples with chicken anti-Zn antibodies, seven main bands with approximate molecular weights 10, 17, 19, 25, 35, 50 and 70 kDa additional bands in variable molecular weights according to sample origin were observed. The intensity and presence of the bands varied within the group. Compared to blots with anti-MT antibody higher background and thus less detectable signals were observed due to interaction of the antibody with blocking agent. However, variable intensities of the bands were determined among individual samples. These bands correspond to zinc-binding proteins and the intensity of the band corresponds to their concentration in the sample.

**Figure 1.** Average levels of MT determined by ELISA and Brdicka reaction and average levels of Zn$^{2+}$ for tumour diseases of interest (medulloblastoma $(n = 10)$, neuroblastoma $(n = 12)$, neuroblastoma HR $(n = 6)$, osteosarcoma $(n = 12)$, Ewing sarcoma $(n = 6)$, Hodgkin lymphoma $(n = 5)$, nephroblastoma $(n = 5)$ and controls $(n = 6)$).
Typical blots stripes of thermally denatured blood sera from patients with neuroblastoma HR, Hodgkin lymphoma, medulloblastoma, osteosarcoma, Ewing sarcoma, nephroblastoma and neuroblastoma are shown in Fig. 2. After visual inspection of both MT and zinc-proteins related bands, the samples exhibit different pattern of bands. At MT blots, the highest variability was determined within the range from 25 to 50 kDa in bands size. However, it is difficult to distinguish the diagnoses visually, especially at nephroblastoma, osteosarcoma and neuroblastoma, where typical bands were missing within the range from 25 to 50 kDa. At Zn blots, the zinc-binding thermostable proteins pattern is not so obvious. On the other hand, it is evident that the protein profiles are not uniform for individual diagnoses. The visual processing of the obtained results clearly showed that there were some associations between diagnoses and bands profile. However, robust mathematical approaches were needed for evaluation of our observations.

3.2 Mathematical processing of the Western blotting images

The samples from the diagnoses sets were randomly distributed within the gels in order to prevent affecting of the mathematical processing results by blotting and electrophoresis conditions. In the typical scanned and treated image, there can be observed numerous bands in the vertical stripes corresponding to different studied samples. Some are clearly visible but some are difficult to be identified by a naked human eye. Western blots fingerprints are often difficult to analyse by common densitometrical software and reproducibility of the profiles are an uneasy task. Automated bioinformatic processing of gel or membrane images enables to mine more information and offers a new tool for proteomic studies making it comparable with mass spectroscopic fingerprint techniques. Mass spectroscopic methods are based on combination of detectors with good resolution and advanced mathematical methods for treatment of data and for comparing them with standard database, which needs to be still updated [61, 62]. That is why we have decided to design an algorithm that would ensure computer processing of the considered image and provide as its output the brightness curves for each of the stripes in the image.

The brightness curve can be further analysed to produce a list of band positions together with some complementary information related to the intensity of the observed bands. Our algorithm starts by separation of individual stripes, continues by generation of a brightness curve for each of the stripes and finally produces an estimate of the molecular weight for each of the identified bands. The brightness curve represents nothing more than aggregated information that is already present in the original image, namely it provides a numeric value corresponding to the intensity of the band (Fig. 2). Clearly, the bands of interest are much easier to be identified in the brightness curves than in the original images because they coincide with the local maxima of the brightness curve (after appropriate smoothing) – compare for example the bands in two lanes of Fig 2. The used algorithmic processing leading to creation of the brightness curves is described in following two paragraphs.

These curves are certainly worth of further attention because their shape seems to carry valuable additional information, which remains to be revealed. We have noticed interesting similarities among shapes of the brightness curves corresponding to certain types of biological samples described in the Section 2.3. Some of them are highlighted below. To support our claim about information hidden in the shape of the considered brightness curves we have used features derived from these graphs as attributes in the classification experiment described in the last paragraph of this section. This experiment proves that the considered features can help in distinguishing between samples of certain tumours, namely embryonal tumours, sarcoma and lymphoma, with significant reliability.
3.2.1 Image segmentation

In the input image, there have to be identified areas of interest (in our particular case these are the vertical stripes). This is the goal of segmentation which ensures that each stripe is analysed separately later. When doing so, we can benefit from strong a priori knowledge we have about electrophoreogram images, for example orientation of the stripes in the vertical direction and a relatively large difference in image brightness values between the areas of interests and the background. That is why we design a dedicated algorithm instead of applying any of the traditional general methods used in image processing (segmentation by luminance, colour or texture, segmentation using edges). Our algorithm segments the image using digital filtering and adaptive thresholding.

3.2.1.1 Segmentation method based on adaptive thresholding

Because the areas of interest in image are vertical stripes of approximately the same width throughout their length, we can simplify the task of image segmentation by turning attention to the sequence of numeric values characterizing the properties of the columns in the image matrix. There are several ways how to provide aggregated information characterising the column, the sum of all values in that column, the average of all the values, their maximum or their minimum. The last approach proved to offer particularly suitable value. Because there is a significant difference between brightness values corresponding to areas of interest and to the background, the sequence of sums of brightness (or of brightness minima) contains information on the location of areas of interest. Under the assumption that the image (when being shot) is evenly lit, it would be possible to obtain image segmentation by simple thresholding. Since this assumption does not apply usually, we do not reach satisfactory segmentation results by this approach. Therefore, it seemed appropriate to apply adaptive thresholding: now, the threshold is somehow modified with respect to the instantaneous values of the thresholded sequence. Thresholding sequence is the first smoothed using suitable low-pass filter to suppress noise and improve success of segmentation. An adaptive threshold is obtained by filtering the sequence by low-pass filter with cut-off frequency lower than that used in the smoothing sequence. The segmentation is obtained simply by comparing the smoothed sequence with the adaptive threshold. The segmentation into individual areas of interest (stripes) is than suggested so that the break points are given by the rising and falling edges in the segmentation thresholding sequence. The described process is shown in Fig. 3A, where we can observe the initial sample sequence of minima characterising individual columns, the corresponding smoothed waveform and its adaptive threshold.

When digital integrator is used as the low-pass filter, this segmentation method is relatively successful in the case of good quality images. The low quality images (that are also present in the treated training set) cannot be reliably segmented in this way. For segmentation of such images we design an alternative method based on autocorrelation function as described later.

3.2.1.2 Digital integrator

Integrator is the easiest filter with infinite impulse response, which has one feedback. It can be represented in block diagram (D1).

In the time domain, the integrator can be described by differential equation:

\[
 y[n] = b_0 \cdot x[n] + a_1 \cdot y[n-1] \tag{1}
\]

Where \( x[n] \) is the sequence to be filtered and \( y[n] \) is the result of its digital filtering. The parameter \( a_0 \) has always value \( a_0 = 1 \). The transfer of this integrator as a function of complex variable \( z \) is then given by the following equation:

\[
 H(z) = \frac{b_0}{a_0 - a_1 \cdot z^{-1}} \tag{2}
\]

3.2.1.3 Segmentation method using the autocorrelation function

This alternative method is very similar to the method described above. Again we filter the input sequence by two different low-pass filters. Instead of comparing the outputs from two low-pass filters as in the case of adaptive filtering, we work with difference of both sequences now. The difference of outputs of two different low-pass filters carries the necessary information for image segmentation, but this sequence itself is not suitable for thresholding. Fortunately, we can rely on the fact that there can be observed periodical changes from areas of interest to the background and back in the image. This periodicity can be identified in the difference sequence, too. Periodicity in sequences can be analysed in various ways including Fourier analysis, the best known method used for that purpose, or autocorrelation. Autocorrelation function is the correlation of the sequence with its shifted image (shift is denoted by \( l \) and its unbiased estimate is defined as follows:

\[
 R_{xx}(l) = \frac{1}{N-1} \sum_{n=1}^{N-l-1} x(n) x(n+l) \tag{3}
\]

While its biased estimate is defined as:

\[
 \hat{R}_{xx}(l) = \frac{1}{N} \sum_{n=1}^{N-l-1} x(n) x(n+l) \tag{4}
\]

Since autocorrelation function is an even function, we can restrict our attention to the estimates of zero and positive translations, only (or to zero and negative translations).
When applying the autocorrelation function for image segmentation, we find out that the positions of local maxima correspond to the areas of interest and local minima to the background in the original image. Adaptive thresholding of the autocorrelation function is demonstrated in the Fig. 3B.

3.2.2 The brightness waveform corresponding to a single stripe

As soon as the individual stripes (or columns) are correctly identified in the original image they have to be further processed to read the information they possess. This information is carried by the pattern of positions where bright horizontal bands (sections) appear. Now, our attention is turned to the individual stripes (columns in the image matrix) identified in the former steps. Individual average brightness waveforms for the considered stripe are extracted as an average over all instances in the area of interest (described by the relevant part of the row in the treated matrix). Using the average brightness waveforms to represent the band proved useful because averaging suppresses the noise that is omnipresent and its suppression is usually suitable. Processing electrophoreogram image to suppress noise proved really necessary because the methods used in the sequel are sensitive to noise. Visual information in the band is very redundant due to the high resolution of images. Thanks to that we do not lose information when representing entire area of interest with average waveform.

3.2.2.1 Finding significant local maxima in the brightness waveforms

The method of searching for local maxima of continuous smooth function can be based on the properties of the first derivative of the function. First derivative of function is positive if the derived function is increasing, and negative value if the function is decreasing. Local extreme (maximum or minimum) is defined as the point where the function changes from increasing to decreasing or from decreasing to increasing. Local extreme in the first derivative of the function corresponds to the zero-crossing. If we want to apply similar principle for finding local extremes in a sequence of discrete values, we need to approximate the derivative. Since our studied sequence corresponds to equidistant measurements, the first derivative can be suitably approximated by the difference between two adjacent values, which can be implemented by digital filtering with differentiator, which is a digital filter with finite impulse response. This process can be described by the differential equation:

\[ y[n] = x[n] - [n - 1] \]  

where \( x[n] \) is the sequence we want to filter and \( y[n] \) is the result of its digital filtering. Transfer of the differentiator as a function of complex variable \( z \) is given by the equation:

\[ H(z) = \frac{z - 1}{z} \]  

Finding zero-crossings in the sequence of the first difference values can be performed by normalizing these values by assigning constant positive value when the difference value is greater than or equal to zero and constant negative value for the difference value is below zero and by repeated calculation of the first difference.

As the search for local maxima sequence uses the first difference, which does not suppress any noise (because it acts as a high-pass filter), we have to remove noise effectively from the brightness waveforms before this method is applied.

To eliminate noise we use the stationary (non-decimated) wavelet transform. We are looking for local maxima in the stationary wavelet approximation of the highest level used in wavelet decomposition. Position of local maxima in the stationary wavelet approximation slightly differs from their exact position in the brightness waveform. To determine the exact position of the local extremes in the brightness waveform the position found in the approximation coefficients has to be recalculated.
All over the wavelet decomposition suppresses noise very successfully, there are identified some local maxima in the result, which do not correspond to bright bands in the considered image, they are erroneous. To decide whether the local maximum is significant or false, we use logical filtering by differences between local maximum and two local minima bordering it. The local maximum is considered as significant if the difference from two neighbouring minima is greater than a specified threshold, or difference from one of the neighbouring minima is greater than the second specified threshold. These two thresholds allow us to distinguish solitary local maxima or those that are superimposed on other significant local maximum. Typical brightness waveform of the standard stripe with labelled local maxima which corresponds to bands is shown in Fig. 3C. Precise identification of these positions is a prerequisite for specification of composition of the considered samples. The automated method described briefly in this paragraph seems to produce very good results but it remains to be verified on some samples of well-known composition, but this topic will be treated in a separate paper.

3.2.2.2 Compact representation of brightness waveform through wavelet transformation

In this paper, we would like to point to the fact that brightness waveforms can play an important role even in classification of the considered samples. To support this argument we will apply the well-known wavelet transformation to the created brightness waveforms with intention to characterise these curves in a concise way using relatively low number of derived aggregated attributes. To do so we have used the wavelet transformation of the level 5 leading to calculation of $2^5$ derived aggregated attributes that provide efficient description of the considered signal by estimating it by a stepwise function (all $2^5$ steps have equal length). This approximation is shown in Figs. 4A and B, for anti-MT and/or anti-Zn blots, respectively. We will show that the wavelet attributes of the considered brightness curves allow designing decision trees that distinguish reasonably well among samples corresponding to various types of tumours. Experiments devoted to design of these decision trees and their testing are described in the following section.

3.2.3 How to classify among various types of tumours using brightness waveform features

Brightness curves of the studied blood samples exhibit interesting similarities (Figs. 3 and 4). This suggests that it might be feasible to use some features characterising these curves as a basis for their classification. To support this claim we have ensured number of experiments during, which decision trees have been constructed from data describing the considered brightness curves by the coefficient of wavelet transformation described in the former paragraph. Due to the used transformation, each curve is described by 32 derived attributes (or coefficients) only. Thus, we are working with a compact representation where each measurements of any blood sample are represented by 32 coefficients. Since each blood sample has been tested by anti-MT and anti-Zn antibodies, each blood sample is represented by $32 + 32 = 64$ coefficients. Our data consist of 56 different blood samples, each is described by $32 + 32$ coefficients and for each there is specified its diagnosis. We intend to design a model that would be able to distinguish among various types of the considered diagnoses using only information about the brightness waveforms of the considered sample. Such models can be designed using methods of machine learning. It is well known that to build a reliable model these methods need to have sufficient amount of data representing typical samples. Moreover, there is a well-known trade-off between complexity of data description,
number of classes into which the data have to be divided and size of the training data. Certainly, much richer set of blood samples would be necessary to build models that would be able to discriminate among all the seven types of considered solid tumours. That is why we have decided to build simpler models that would provide less detailed characterisation of the considered tumours (abstracting, e.g. from their precise localisation as in neuroblastoma): our models distinguish among embryonal tumours, lymphoma and sarcoma in general only. Even this task is very difficult provided that we have no more than 56 samples in total, each described by 32 + 32 attributes. We have decided to apply the well-known algorithm for decision tree building to design a model of the way how the diagnosis can be based on the values of coefficients characterising our data.

The algorithm for building a decision tree model from the training data applies iteratively a routine for identification of the most informed attribute to the considered dataset. This attribute is then used for partitioning the dataset. Further on significantly smaller dataset is processed and complexity of the considered task is reduced in this way step by step. Those attributes that appear in the names of the upper nodes of the decision tree (close to its root) seem to be of special importance and the decision tree can be understood as a feature selection algorithm, too. The software tool Rapid Miner and its module Decision Tree have been applied.

Classification models have been designed separately in two experiments:

(i) All samples have been considered and the models have been designed to classify them into three classes: embryonal tumours, lymphoma and sarcoma. Here, there is a danger of misclassification in case of lymphoma, since this class is significantly underrepresented in the original data consisting of 28 samples of embryonal tumours, 23 samples of sarcoma and 5 samples of lymphoma only.

(ii) The five lymphoma samples have been neglected and the classification task was studied using only 51 samples that have to be separated into two classes: embryonal tumours and sarcoma.

To test the quality of the created models and to compare results of both experiments there was applied ten-fold cross-validation in both cases. This means that the available data S corresponding to different brightness waveforms were divided into ten disjunctive sets \( S_1, \ldots, S_{10} \) each of which maintained the same percentage of the considered body tissues as the original set (stratified samples). For \( i = 1 \) to 10 the following experiments have been ensured: the decision tree model has been created from the training data set \( S_i \) and tested on the remaining data, namely on \( S_i \). The overall results of the ten experiments were summarised using a confusion matrix with columns denoted by the type of tumour as diagnosed by the medical expert and with rows denoted by the classification suggested by the constructed decision trees. Consequently, the resulting table denoted as a confusion matrix for the considered experiment depicts all the correctly classified examples on the diagonal of the matrix while all the other points represent errors. Moreover, this presentation of the obtained results makes it possible to identify easily the most frequent mistakes or confusions appearing in our data, namely the names of solid tumour type, the miss-classified examples came from and the predicted class, which does not seem to be fully reliable.

The obtained results summarised in Fig. 5, and Table 1 prove that the suggested approach for using the wavelet coefficients for tumour diagnostics is very promising. After analysis of the blots there was possible to classify individual path according to tumour (embryonal tumours, sarcoma, lymphoma). In Zn blots, the accuracy was 88% for embryonal tumours and 69% for sarcomas, while we were not able to classify lymphomas. The accuracy for MT blots was
Table 1. Results of ten-fold crossvalidation for decision tree building, Zn–embryonal tumours, sarcoma, lymphoma

<table>
<thead>
<tr>
<th></th>
<th>True embryonal</th>
<th>True lymphoma</th>
<th>True sarcoma</th>
<th>Class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred. blastoma</td>
<td>44</td>
<td>4</td>
<td>7</td>
<td>80.00%</td>
</tr>
<tr>
<td>Pred. lymphoma</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0.00%</td>
</tr>
<tr>
<td>Pred. sarcoma</td>
<td>6</td>
<td>5</td>
<td>24</td>
<td>68.57%</td>
</tr>
<tr>
<td>Class recall</td>
<td>88.00%</td>
<td>0.00%</td>
<td>68.57%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>71.58 ± 13.77%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Results of ten-fold crossvalidation for decision tree building, MT – embryonal tumours, sarcoma, lymphoma

<table>
<thead>
<tr>
<th></th>
<th>True embryonal</th>
<th>True lymphoma</th>
<th>True sarcoma</th>
<th>Class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred. blastoma</td>
<td>21</td>
<td>3</td>
<td>8</td>
<td>65.62%</td>
</tr>
<tr>
<td>Pred. lymphoma</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>25.00%</td>
</tr>
<tr>
<td>Pred. sarcoma</td>
<td>9</td>
<td>1</td>
<td>11</td>
<td>52.38%</td>
</tr>
<tr>
<td>Class recall</td>
<td>65.62%</td>
<td>20.00%</td>
<td>55.00%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>58.33 ± 20.72%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Results of ten-fold crossvalidation for decision tree building, Zn + MT – embryonal tumours, sarcoma, lymphoma

<table>
<thead>
<tr>
<th></th>
<th>True embryonal</th>
<th>True lymphoma</th>
<th>True sarcoma</th>
<th>Class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred. blastoma</td>
<td>21</td>
<td>1</td>
<td>5</td>
<td>77.78%</td>
</tr>
<tr>
<td>Pred. lymphoma</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>14.29%</td>
</tr>
<tr>
<td>Pred. sarcoma</td>
<td>0</td>
<td>3</td>
<td>11</td>
<td>78.57%</td>
</tr>
<tr>
<td>Class recall</td>
<td>84.00%</td>
<td>20.00%</td>
<td>61.11%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>69.00 ± 22.00%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Results of ten-fold crossvalidation for decision tree building, Zn – embryonal tumours, sarcoma

<table>
<thead>
<tr>
<th></th>
<th>True embryonal</th>
<th>True lymphoma</th>
<th>True sarcoma</th>
<th>Class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred. blastoma</td>
<td>23</td>
<td>0</td>
<td>4</td>
<td>85.19%</td>
</tr>
<tr>
<td>Pred. lymphoma</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Pred. sarcoma</td>
<td>3</td>
<td>0</td>
<td>14</td>
<td>82.35%</td>
</tr>
<tr>
<td>Class recall</td>
<td>88.46%</td>
<td>0.00%</td>
<td>77.78%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>82.33 ± 22.41%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Results of ten-fold crossvalidation for decision tree building, MT – embryonal tumours, sarcoma

<table>
<thead>
<tr>
<th></th>
<th>True embryonal</th>
<th>True lymphoma</th>
<th>True sarcoma</th>
<th>Class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred. blastoma</td>
<td>17</td>
<td>0</td>
<td>11</td>
<td>60.71%</td>
</tr>
<tr>
<td>Pred. lymphoma</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Pred. sarcoma</td>
<td>15</td>
<td>0</td>
<td>9</td>
<td>37.50%</td>
</tr>
<tr>
<td>Class recall</td>
<td>53.12%</td>
<td>0.00%</td>
<td>45.00%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>49.83 ± 25.10%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Results of ten-fold crossvalidation for decision tree building, Zn + MT – embryonal tumours, sarcoma

<table>
<thead>
<tr>
<th></th>
<th>True embryonal</th>
<th>True lymphoma</th>
<th>True sarcoma</th>
<th>Class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred. blastoma</td>
<td>20</td>
<td>0</td>
<td>6</td>
<td>76.92%</td>
</tr>
<tr>
<td>Pred. lymphoma</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Pred. sarcoma</td>
<td>5</td>
<td>0</td>
<td>12</td>
<td>70.59%</td>
</tr>
<tr>
<td>Class recall</td>
<td>80.00%</td>
<td>0.00%</td>
<td>66.67%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>76.00 ± 23.85%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

66% for embryonal tumours, 20% for lymphomas and 55% for sarcomas (Table 2). When analysed Zn and MT blots both accuracies increased to 84% for embryonal tumours and was 20% for lymphomas and 61% for sarcomas (Table 3). Because of low number of lymphomas we tested only classification of embryonal tumours. In Zn blots, the accuracy was 88% for embryonal tumours and 78% for sarcomas (Table 4). In MT blots, the accuracy was 53% for embryonal tumours and 45% for sarcomas (Table 5). After analysing both blots together the accuracy was 80% for embryonal tumours and 67% for sarcomas (Table 6). It is interesting that although we used heat-treated samples, where the majority representation of MT is assumed [16,54], there were determined enough other thermostable Zn containing proteins in samples. Those proteins served to identify embryonal tumours with accuracy 88% and sarcomas with 78% accuracy, while in MT blots, the accuracy were only 53% in embryonal tumours and 45% in sarcomas. Analysis both Zn and MT blots simultaneously did not increase accuracy of identification neither in embryonal tumours (80%) nor in sarcomas (67%). We do not know which of thermostable zinc-containing proteins enable to diagnose those tumours. Further biochemical studies will be necessary to clarify importance those proteins for diagnosis and pathogenesis of malignant proteins.

4 Concluding remarks

We found differences both in Zn and MT levels in childhood solid tumours compared to healthy controls. The results were subjected bioinformatic processing enabled on a different pattern of bands to distinguish the groups of embryonal tumours and sarcomas. Those results are promising not only from diagnostic point of view but particularly in the area of basic research. Using these data, we can be able to further study individual MT isoforms and their aggregated forms in malignant tumours. Important was also possibility to test bioinformatics processing on real data.
The authors have declared no conflict of interest.

5 References