

The use of infrared spectroscopy and artificial neural networks for detection of uropathogenic *Escherichia coli* strains' susceptibility to cephalothin*

Łukasz Lechowicz¹✉, Mariusz Urbaniak², Wioletta Adamus-Białek³ and Wiesław Kaca¹

¹Department of Microbiology, Jan Kochanowski University, Kielce, Poland; ²Organic Chemistry Division, Jan Kochanowski University, Kielce, Poland; ³Independent Department of Environmental Protection and Modeling, Jan Kochanowski University, Kielce, Poland

Background & Aims: Infrared spectroscopy is an increasingly common method for bacterial strains' testing. For the analysis of bacterial IR spectra, advanced mathematical methods such as artificial neural networks must be used. The combination of these two methods has been used previously to analyze taxonomic affiliation of bacteria. The aim of this study was the classification of *Escherichia coli* strains in terms of susceptibility/resistance to cephalothin on the basis of their infrared spectra. The infrared spectra of 109 uropathogenic *E. coli* strains were measured. These data are used for classification of *E. coli* strains by using designed artificial neural networks. **Results:** The most efficient artificial neural networks classify the *E. coli* sensitive/resistant strains with an error of 5%. **Conclusions:** Bacteria can be classified in terms of their antibiotic susceptibility by using infrared spectroscopy and artificial neural networks.

Key words: infrared spectroscopy, artificial neural network, uropathogenic *Escherichia coli*, antibiotic resistance

Received: 31 October, 2013; **revised:** 12 December, 2013; **accepted:** 20 December, 2013; **available on-line:** 30 December, 2013

INTRODUCTION

Multi-drug resistant strains are one of the main problems in the bacterial infection treatment. Cephalothin is a semisynthetic antibiotic belonging to the first generation of cephalosporins. It is used as a therapeutic agent in infections caused by Gram-negative bacteria - *Escherichia coli*, *Proteus mirabilis*, *Klebsiella pneumoniae* (Empel *et al.*, 2008). Cephalosporins, including cephalothin, are inactivated by ESBL β -lactamases i.e. AmpC, TEM, SVH or MBL (Prabhu *et al.*, 2013). Percentage of uropathogenic *E. coli* strains resistant to first class cephalosporins is still increasing (Kot *et al.*, 2010). Drug-resistant strains are isolated from environmental samples as well as from clinical samples.

There are several techniques for determining the susceptibility of bacteria to antibiotics. Most of the techniques (e.g. disc diffusion method and E-test) are based on the cultivation of bacteria in the presence of an antibiotic. It is also possible to detect genes for resistance to antibiotics by using PCR (Empel *et al.*, 2008). These methods, however, require time and are labor intensive.

Traditional methods of bacterial phenotypic assessment i.e. antibiotics resistance/sensitivities are based on growing bacterial cultures on specific media (e.g. Christensen medium). Strains are differentiated on the basis of the appearance of bacterial colonies and micro-

bial medium changes. Modern chemical methods allow the phenotypic differentiation of bacteria based on the presence of specific peaks in bacterial spectrum reflecting presence of biopolymers of the intact bacterial cells. One of such methods is the Fourier Transform Infrared Spectroscopy (FTIR) (Lechowicz *et al.*, 2013). On the basis of FTIR technique, reflection techniques have been developed which allow obtaining an IR spectrum by measuring light reflected by the sample. The most commonly used method is the Attenuated Total Reflection (ATR/FTIR). The ATR/FTIR technique has made the infrared spectroscopy a powerful tool for the analysis of systems previously inaccessible for it. Fast, cheap, non-invasive method that does not require special preparation of samples quickly found use in biological research (Maquelin *et al.*, 2002; Yu *et al.*, 2005). The superposition of signals from many biomolecules present in a bacterial cell makes peak assignments difficult, but highly specific patterns of IR spectra of intact bacterial cells may be employed for the rapid and accurate identification of microorganisms and their phenotypic characteristics. This method has distinguished five major fragments of spectrum useful for the bacterial identification: 3000–2800 cm^{-1} fragment (or W1 window) is the fatty acid region; 1700–1500 cm^{-1} (W2 window) contains the amide I and II bands of proteins and peptides; 1500–1200 cm^{-1} (W3 window) is a mixed region of fatty acid bending vibrations, proteins, and phosphate-carrying compounds; 1200–900 cm^{-1} (W4 window) contains absorption bands of the carbohydrates; and 900–700 cm^{-1} (W5 window) is the 'fingerprint region' that contains unique peaks that are characteristic for specific species of bacteria (Fig. 1) (Helm *et al.*, 1991a; Helm *et al.*, 1991b; Naumann *et al.*, 1991). Infrared spectra of intact microbial cells are very complex. Attempts have been made to use mathematical methods in the analysis of IR spectra of biological samples. The most popular mathematical methods include principal component analysis (PCA), cluster analysis (CA) and artificial neural networks (ANNs) (Mouwen *et al.*, 2006; Bosch *et al.*, 2008; Wenning *et al.*, 2010; Lechowicz *et al.*, 2013). Artificial neural networks are mathematical models of the nervous system. The basic units of ANNs are artificial neurons. Construction of artificial

✉ e-mail: lechowiczlukasz@gmail.com

*Part of the results was presented at the 3-rd Workshop on Microbiology „MIKROBIOT 2013” in Łódź, Poland.

Abbreviations: ANNs, artificial neural networks; ATR, attenuated total reflection; CA, cluster analysis; FTIR, fourier transform infrared spectroscopy; IR, infrared; MLP, multilayer perceptron; PCA, principal component analysis; RBF, radial basis function networks

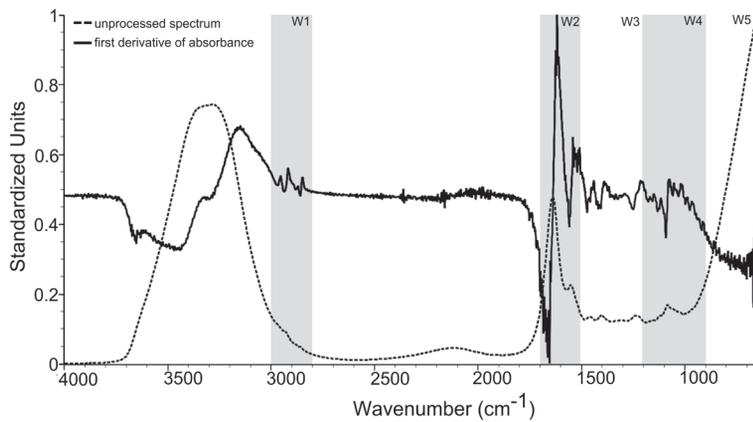


Figure 1. An example of infrared spectrum of uropathogenic *E. coli* strain.

neurons is analogous to the construction of a biological neuron (Fig. 2a). There are several types of neural networks, differing in arrangement of artificial neurons. Frequently used ANNs are composed of one neuron layer for data input, two or three layers of hidden neurons and a neuron layer leading-out data from the network (Fig. 2b). ANNs have found many applications in the analysis of various types of data. The learning process of ANN is based on presentation of the data (learning set) to network. After learning, the network optimizes

its structure and thereby increases the number of correct classifications. The main limitation of this technique is the need to have a large number of learning data. An increase in the complexity of the network makes it necessary to increase the size of the learning set. To solve very complex problems, very complex neuronal networks need to be applied and thus the learning set must count thousands of observations.

The aim of this study was the use of IR spectroscopy and artificial neural networks to differentiate cephalothin (cephalosporine) resistant/sensitive uropathogenic *E. coli* strains.

MATERIALS AND METHODS

The bacterial strains. 109 uropathogenic *E. coli* strains were isolated from urine samples of hospitalized patients suffering from urinary tract infection, from the wards of hospital No. 2, Medical University of Lodz, Poland, in the years 2006 to 2008. Following isolation, the strains were stored at -80°C until the tests were carried out. Number of passages between the isolation and performance of the test was not more than five. Bacteria were grown at 37°C for 24 hours. After this time the bacteria were stored at 4°C .

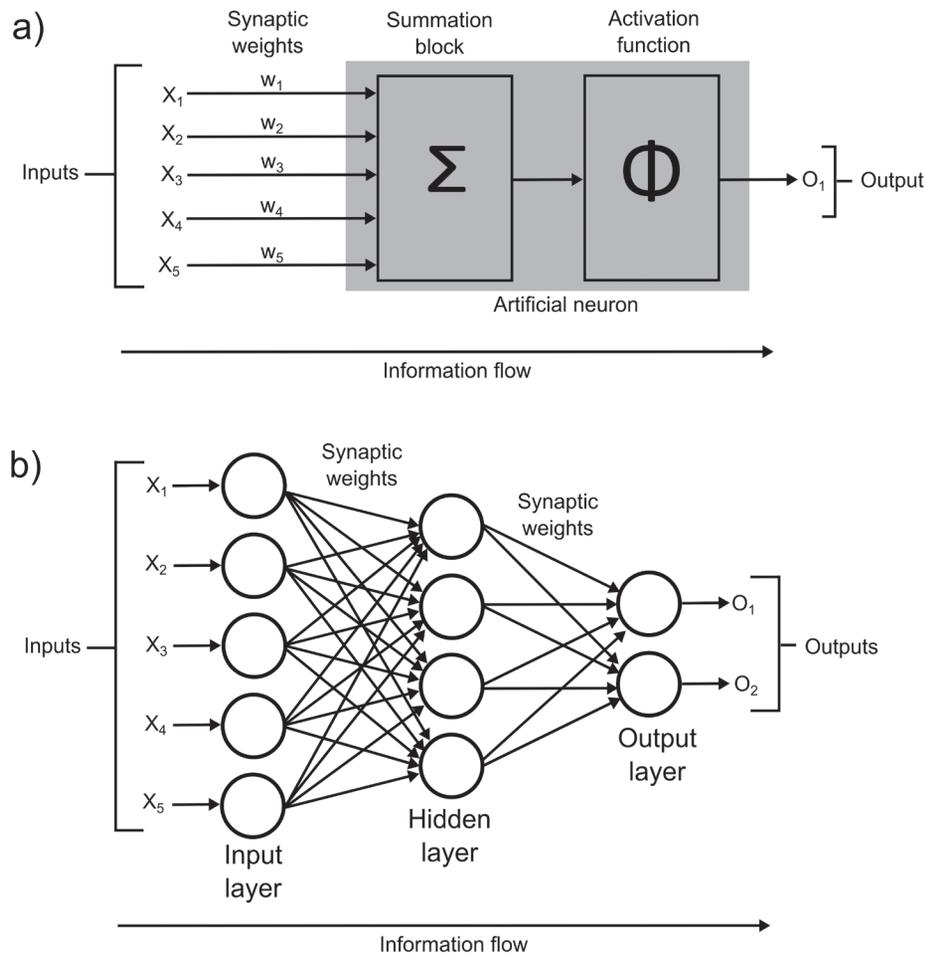


Figure 2. Construction of an artificial neuron (a) and the structure of an artificial neural network (b).

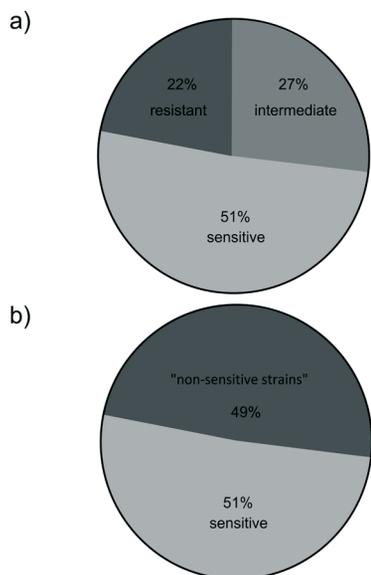


Figure 3. The percentage of *E. coli* strains susceptible to cephalothin used in the study.

(a) distribution of strains into three categories — resistant, sensitive and intermediate (b) distribution of strains into two categories — sensitive and "non-sensitive strains"

Designation of drug resistant strains. The drug resistance of *E. coli* strains was determined by a susceptibility test (ATB UR5, bioMerieux) at the Faculty of Laboratory Diagnostics and Clinical Biochemistry, hospital No. 2, Medical University of Lodz, Poland (Adamus-Bialek W *et al.*, 2009; Adamus-Bialek W *et al.*, 2013).

Measurement of IR spectra. Bacterial strains were cultured on LB agar medium at 37°C for 24 hours. A single bacterial colony was very carefully taken from

the Petri dish using a disposable sterile pipette tip, just before the IR spectrum measurement. The bacterial mass was evenly distributed over the entire surface of the crystal. Bacterial colony was scanned 25 times, and the final result was averaged. For each bacterial strain, 10 colonies were harvested. After IR spectrum measurement, the bacterial mass was removed using acetone and 70% isopropanol. Residual solvents were removed with a sterile double-distilled water, and the crystal was dried. IR spectrum was measured in the range 4000–650 cm^{-1} with an interval of 1 cm^{-1} . The resulting spectrum of bacteria was then subjected to a pre-treatment: the baseline fixed, calculation of the first derivative, scaling to the range 0–1.

Designing a neural network. For creating a neural network Statistica 10 was used. The calculations were based on data from 1090 bacterial IR spectra. The IR spectra were randomly divided into independent subsets: learning subset (70% of all IR spectra), testing subset (15% of all IR spectra) and validation subset (15% of all IR spectra).

RESULTS

Distribution of bacterial strains due to the susceptibility to cephalothin was as follows: 51% of the bacterial strains were sensitive to cephalothin, 27% of the bacterial strains were resistant to cephalothin, and the remaining 22% were displaying intermediate resistance to cephalothin (Fig. 3a) (Adamus-Bialek *et al.*, 2009; Adamus-Bialek *et al.*, 2013). Unprocessed IR spectra of different bacterial strains were similar. However, after conversion of original spectra to the first derivative some differences could be observed. Especially large accumulation of the differences may be observed in the fragment 1500–1000 cm^{-1} which corresponds to windows W3, W4 (Fig. 4a, b). Nonetheless these differences are so slight and unstable that the differentiation of resistant/sensitive of *E. coli* on their basis is not possible. The wavenumbers related to cephalothin resistance have been identified by using statistical methods. Advanced methods of multivariate data processing (eg. PCA, CA) did not allow to differentiate the *E. coli* strains (data not shown). We designed a set of artificial neural network to classify the tested bacteria based on the selected wavenumbers. In this study two types of neural networks were used: Multilayer Perceptron (MLP) and Radial Basis Function networks (RBF). The individual networks differed in the number of neurons in the hidden layer and the type of the activation function. The quality of the network was evaluated on the basis of the number of correct classification in the validation subset, and in the second place, the network complexity. The best results were obtained with the MLP networks (Table 1). The performance of RBF networks was significantly below expectations (Table 2). The best designed ANN obtained 83.43% of correctly classified strains (or correct answers) in the validating set (Table 1). This result is decidedly unsatisfactory. However, after examining the responses distribution in the validation

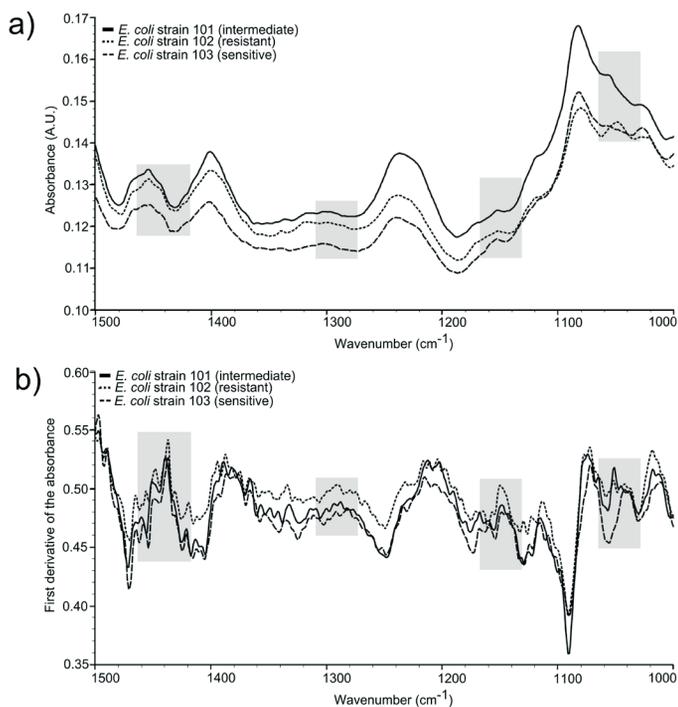


Figure 4. Differences in the IR spectra of three uropathogenic *E. coli* strains.

(a) unprocessed spectra (b) the first derivative of absorbance. Some differences are marked by grey areas.

Table 1. Multilayer perceptrons (MLP) for classifying bacterial strains into 3 groups.

ID	Topology	Quality of the network			Error function	Activation function	
		Training (%)	Testing (%)	Validating (%)		Hidden neurons	Output neurons
1	MLP 30-8-3	89.66	79.14	83.43	Ent	Log	Softmax
2	MLP 30-10-3	94.11	80.37	79.75	Ent	Log	Softmax
3	MLP 30-6-3	84.95	77.91	79.14	Ent	Tanh	Softmax
4	MLP 30-9-3	89.66	77.30	78.53	Ent	Log	Softmax
5	MLP 30-9-3	91.75	79.14	77.91	Ent	Log	Softmax
6	MLP 30-8-3	86.78	74.85	77.91	Ent	Tanh	Softmax
7	MLP 30-9-3	89.66	79.14	77.30	Ent	Tanh	Softmax
8	MLP 30-7-3	90.84	77.91	77.30	Ent	Log	Softmax
9	MLP 30-9-3	81.94	73.00	76.69	Ent	Log	Softmax
10	MLP 30-6-3	85.73	78.53	76.69	Ent	Tanh	Softmax

Abbreviations: Tanh — hyperbolic tangent, Log — logistic function, Ent — entropy, Softmax — softmax function.

Table 2. Radial basis function networks (RBF) for classifying bacterial strains into 3 groups.

ID	Topology	Quality of the network			Error function	Activation function	
		Training (%)	Testing (%)	Validating (%)		Hidden neurons	Output neurons
1	RBF 30-10-3	52.75	60.12	58.90	SOS	Gaussian	Linear
2	RBF 30-10-3	53.01	58.28	58.28	Ent	Gaussian	Softmax
3	RBF 30-10-3	52.88	60.12	57.67	SOS	Gaussian	Linear
4	RBF 30-10-3	54.19	58.90	57.67	SOS	Gaussian	Linear
5	RBF 30-10-3	56.68	56.44	57.06	Ent	Gaussian	Softmax
6	RBF 30-10-3	51.70	56.44	56.44	Ent	Gaussian	Softmax
7	RBF 30-10-3	50.92	60.74	55.83	Ent	Gaussian	Softmax
8	RBF 30-10-3	52.75	55.83	55.83	Ent	Gaussian	Softmax
9	RBF 30-10-3	53.40	58.90	55.83	SOS	Gaussian	Linear
10	RBF 30-10-3	53.27	59.51	55.21	SOS	Gaussian	Linear

Abbreviations: Gaussian — Gaussian function, Softmax — softmax function, Linear — linear function, SOS — sum of squares, Ent — entropy.

Table 3. Compatibility of the results obtained from artificial neural network (see Table 1. ID 1) used for *E. coli* classification with cephalothin susceptibility testing for the validation subset.

	Cephalothin susceptibility testing		
	Resistant strains	Intermediate strains	Sensitive strains
Correctly classified strains	67.5 %	77 %	95 %
Incorrectly classified strains	32.5 %	23 %	5 %

set it turns out that the percentage of correctly classified strains for cephalothin-sensitive strains was about 95%. For resistant and intermediate strains, the ANN obtained significantly worse results — 67.5% and 77% of correctly classified strains, respectively (Table 3). It is a well-known phenomenon of ANN functioning. Sometimes ANNs demonstrate a capacity to recognize one type of pattern (e.g. sensitive strains) and at the same time do not recognize the other patterns (e.g. intermediate and resistant strains). This particular network showed some predisposition for detecting susceptibility to cephalothin. In the first experiment, networks were designed for classifying bacterial strains into three types of antibiotic action — resistant, intermediate and sensitive. However, ANNs performed better when only two types of

strains were analyzed. Therefore, in the next experiment, intermediate and resistant strains were combined into one group called “non-sensitive strains”. The new distribution of strains is shown in Fig. 3b. In this case, the topology of the ANNs was changed — networks were constructed with 30 input neurons, 9 neurons in the hidden layer and two outputs. For data prepared this way, neural networks obtained slightly different results. Also, in this case MLP networks achieved the best results (Table 4). In the validation set, the best network obtained 92.64% of correctly classified strains. For cephalothin-sensitive strains the number of correct answers was 93.23%. It is therefore a slightly worse result than that obtained in the first experiment. However, for strains grouped in the category “non-sensitive strains” percent-

Table 4. Multilayer perceptrons (MLP) for classifying bacterial strains into 2 groups.

ID	Topology	Quality of the network			Error function	Activation function	
		Training (%)	Testing (%)	Validating (%)		Hidden neurons	Output neurons
1	MLP 30-9-2	98.82	90.80	92.64	SOS	Log	Linear
2	MLP 30-9-2	98.04	92.64	92.02	SOS	Tanh	Tanh
3	MLP 30-9-2	95.29	92.02	91.41	SOS	Tanh	Tanh
4	MLP 30-8-2	97.38	92.64	91.41	SOS	Log	Linear
5	MLP 30-10-2	97.91	90.80	91.41	SOS	Tanh	Linear
6	MLP 30-8-2	97.25	90.80	90.80	SOS	Log	Tanh
7	MLP 30-8-2	97.51	92.02	90.80	SOS	Tanh	Tanh
8	MLP 30-10-2	96.99	90.80	90.80	SOS	Log	Linear
9	MLP 30-10-2	98.69	90.80	90.80	SOS	Tanh	Linear
10	MLP 30-10-2	95.81	88.96	90.80	SOS	Exp	Linear

Abbreviations: Tanh — hyperbolic tangent, Log — logistic function, Exp — exponential function, Linear — linear function, SOS — sum of squares.

Table 5. Radial basis function networks (RBF) for classifying bacterial strains into 2 groups.

ID	Topology	Quality of the network			Error function	Activation function	
		Training (%)	Testing (%)	Validating (%)		Hidden neurons	Output neurons
1	RBF 30-3-2	59.82	65.64	65.03	Ent	Gaussa	Softmax
2	RBF 30-3-2	63.22	63.80	65.03	Ent	Gaussa	Softmax
3	RBF 30-3-2	61.52	65.64	64.42	SOS	Gaussa	Liniowa
4	RBF 30-3-2	60.21	61.96	64.42	Ent	Gaussa	Softmax
5	RBF 30-3-2	58.12	57.06	64.42	Ent	Gaussa	Softmax
6	RBF 30-3-2	59.95	57.67	64.42	Ent	Gaussa	Softmax
7	RBF 30-30-2	57.98	58.90	63.80	Ent	Gaussa	Softmax
8	RBF 30-3-2	57.07	57.67	63.80	Ent	Gaussa	Softmax
9	RBF 30-3-2	65.18	65.64	63.80	SOS	Gaussa	Linear
10	RBF 30-3-2	59.42	60.74	63.80	SOS	Gaussa	Linear

Abbreviations: Gaussa — Gaussian function, Softmax — softmax function, Linear — linear function, SOS — sum of squares, Ent — entropy.

age of correct classification was 91.9%, which should be considered as a good result. Also in this case, the RBF network obtained validation of 64% which was not satisfactory (Table 5).

DISCUSSION

One of the main problems in microbial diagnostics is labor and time-consuming drug-resistance testing. In case of a life-threatening bacterial infection (e.g. during systemic inflammatory response) it is necessary to ensure proper antibiotic treatment within a few hours of onset of the symptoms. Through the use of modern chemometric methods it will be possible in the future to rapidly determine drug susceptibility of bacteria within several hours. Among such methods should be mentioned FTIR and MALDI-TOF. So far, there have been reports about the possibility of using IR spectroscopy and PCA in the study of imipenem-resistance of *Pseudomonas* sp (Sokalingum *et al.*, 1997). No attempt was made to use artificial neural networks in such studies. A method of storing information of drug resistance in the IR spectrum

of bacteria is not clear. There are several mechanisms for antibiotic resistance: the presence of specific enzymes (e.g. β -lactamase), transmembrane proteins and others. Possibly, the presence of these cellular components is reflected in the IR spectrum of bacteria. These characteristics are not concentrated in a single fragment of the spectrum (or one peak), but dispersed over the whole length of the spectrum. *E. coli* strains may differ from each other in terms of genetic and metabolic characteristics. This makes the problem of drug-resistant strain identification very complex. The use of mathematical methods (e.g. CA, PCA) may not provide the expected validation and correct results. Differences in the phenotype of *E. coli* can cause distortion of the IR spectrum and obliteration of search features. It can be compared in the presence of random noise in the IR spectrum. One of the characteristics of neural networks is their resistance to certain distortions of data. This makes it possible to better recognize the patterns in the analyzed data. Based on these characteristics it can be concluded that the ANNs are one of the methods of choice for the analysis of the bacterial IR spectra.

CONCLUSIONS

The combination of infrared spectroscopy (ATR/FTIR) of bacterial samples and advanced methods of data analysis (ANNs) may allow rapid differentiation of bacterial strains in terms of their antibiotic resistance/sensitivity.

Acknowledgements

This project was financed from the funds of the National Science Center allocated on the basis of the decision number DEC-2012/07/N/NZ7/01187.

REFERENCES

- Adamus-Bialek W, Wojtasik A, Majchrzak M, Sosnowski M, Parniewski P (2009) (CGG)4-Based PCR as a novel tool for discrimination of uropathogenic *Escherichia coli* Strains: comparison with enterobacterial repetitive intergenic consensus-PCR. *Journal of Clinical Microbiology* **47**: 3937–3944.
- Adamus-Bialek W, Zajac E, Parniewski P, Kaca W (2013) Comparison of antibiotic resistance patterns in collections of *Escherichia coli* and *Proteus mirabilis* uropathogenic strains. *Mol Biol Rep* **40**: 3429–3435.
- Bosch A, Minan A, Vescina C, Degrossi J, Gatti B, Montanaro P, Messina M, Franco M, Vay C, Schmitt J, Naumann D, Yantorno O (2008) Fourier Transform Infrared Spectroscopy for rapid identification of nonfermenting gram-negative bacteria isolated from sputum samples from cystic fibrosis patients. *J Clin Microbiol* **46**: 2535–2546.
- Empel J, Baraniak A, Literacka E, Mrówka A, Fiett J, Sadowy E, Hryniewicz W, Gniadkowski M; Beta-PL Study Group. (2008) Molecular survey of beta-lactamases conferring resistance to newer beta-lactams in *Enterobacteriaceae* isolates from Polish hospitals. *Antimicrob Agents Chemother* **52**: 2449–2454.
- Helm D, Labischinski H, Naumann D (1991) Elaboration of a procedure for identification of bacteria using Fourier Transform IR spectral libraries: a stepwise correlation approach. *J Microbiol Methods* **14**: 127–142.
- Helm D, Labischinski H, Schallehn G, Naumann D (1991) Classification and identification of bacteria by Fourier-transform infrared spectroscopy. *J Gen Microbiol* **137**: 69–79.
- Kot B, Wicha J, Zak-Pulawska Z (2010) Susceptibility of *Escherichia coli* strains isolated from persons with urinary tract infections in 2007–2008 to antimicrobial agents. *Przegl Epidemiol* **64**: 307–312.
- Lechowicz L, Adamus-Bialek W, Kaca W (2013) Attenuated Total Reflectance Fourier Transform Infrared spectroscopy and artificial neural networks applied to differentiate *Escherichia coli* papG+/papG-strains. *J Spectro* **2013**: 538686 doi:10.1155/2013/538686.
- Maquelin K, Kirschner C, Choo-Smith LP, van den Braak N, Endtz HP, Naumann D, Puppels GJ (2002) Identification of medically relevant microorganisms by vibrational spectroscopy. *J Microbiol Methods* **51**: 255–271.
- Mouwen DJM, Capita R, Alonso-Calleja C, Prieto-Gómez J, Prieto M (2006) Artificial neural network based identification of *Campylobacter* species by Fourier transform infrared spectroscopy. *J Microbiol Methods* **67**: 131–140.
- Naumann D, Helm D, Labischinski H (1991) Microbiological characterizations by FT-IR spectroscopy. *Nature* **351**: 81–82.
- Prabhu A, Taylor P, Konecny P, Brown MA (2013) Pyelonephritis: what are the present day causative organisms and antibiotic susceptibilities? *Nephrology* **18**: 463–467.
- Sockalingum DG, Bouhedja W, Pina P, Allouch P, Mandray C, Labia R, Millot JM, Manfait M (1997) ATR–FTIR spectroscopic investigation of imipenem-susceptible and -resistant *Pseudomonas aeruginosa* isogenic strains. *Biochem Biophys Res Commun* **232**: 240–246.
- Wenning M, Büchl NR, Scherer S (2010) Species and strain identification of lactic acid bacteria using FTIR spectroscopy and artificial neural networks. *J Biophotonics* **3**: 493–505.
- Yu C, Irudayaraj J (2005) Spectroscopic characterization of microorganisms by Fourier transform infrared microspectroscopy. *Biopolymers* **77**: 368–377.