pattern recognition, fuzzy k-NN classifier, ROC analysis,
DNA repair, adaptive response, mutagenesis,
AlkB dioxygenase, AlkA glycosylase, E.coli

Beata SOKOŁOWSKA[1], Agnieszka M. MACIEJEWSKA[1], Adam JÓŹWIK[2],
Jarosław T. KUŚMIEREK[1]

ESTIMATION OF SIGNIFICANCE OF ALKB AND ALKA PROTEINS IN DNA REPAIR IN *ESCHERICHIA COLI* MODEL

The paper concerns estimation of significance of differences of mutagenesis level between the wild-type strain (*wt*) and its derivatives which differ in DNA repair ability, namely *alkA* and *alkB* strain, devoided AlkA glycosylase and AlkB dioxygenase activity, respectively. The strains were analyzed for their ability to repair 1,N^6-ethenoadenine (εA) – chloroacetaldehyde adduct to DNA. The analysis was done using classical statistical and pattern recognition methods. The obtained results confirmed that AlkB dioxygenase plays the most important role in εA repair in *E. coli* in the experimental modeling.

1. INTRODUCTION

DNA is subjected to endogenous and exogenous influences which may affect its structure and integrity [1,2]. Damages may be induced by several chemical reactive species or physical and biological agents. Many of the lesions occur spontaneously through intrinsic instability of DNA bonds. At the organism level, DNA lesions are implicated in various inherited and genetic diseases [10,24], in carcinogenesis [14], and also in aging [19]. The cells have developed a network of several complementary DNA repair mechanisms [1]. The adaptive response system (Ada response) involves *ada, alkA, alkB* and *aidB* genes in *E.coli* [8,27]. These genes are expressed after exposure to non-toxic doses of direct-acting alkylating agents [13,25,26]. In experimental conditions chloroacetaldehyde (i.e. CAA, $ClCH_2CHO$), which is a metabolite of carcinogen vinyl chloride, can be used to generate etheno (ε) adducts in DNA, such as 3,N^4-ethenocytosine (εC), 1,N^6-ethenoadenine, N^2,3-ethenoguanine (N^2,3εG) and 1,N^2-ethenoguanine (1,N^2εG) [4,15,21]. These exocyclic DNA adducts are also generated endogenously as products of peroxidation of membrane lipids [3, 22]. For example, AlkA glycosylase (AlkA) removes N^2,3εG whereas 3,N^4εC and 1,N^6εA are substrates for AlkB dioxygenase (AlkB) [6,17,18,20].

The experimental data were obtained using a system developed in our laboratory. We have used pIF105 plasmid [11] carrying *lacZ* allele from CC105 strain [5], indicatory for AT→TA base substitutions, main type of mutations caused by εA [6]. The plasmid DNA was modified with CAA *in vitro,* and then DNA repair processes of the plasmid were studied *in vivo* in *E.coli* cells. The use of wild-type *E.coli* and its mutants either in AlkB dioxygenase as well as in AlkA glycosylase allowed estimating the role of each enzyme in repair of εA in bacterial cells. Induction of adaptive response to alkylating agents increases the level of AlkA and AlkB proteins in the bacterial cells [23, 26].

The system was previously successfully applied to study CAA-induced GC→AT (for pIF102 plasmid) and GC→TA (for pIF104 plasmid) mutations and to estimate the role of AlkB in repair of cytosine-CAA adducts [16,17].

Review of literature data points out that two proteins of Ada response, AlkB and AlkA are involved in repair of ethenoadducts. We have suggested that AlkB protein is more active than AlkA in εA repair [18]. The aim of present work is to verify the hypothesis using a complex statistical analysis of the experimental data obtained from mutagenesis test.

[1] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland.
[2] Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland.

2. MATERIALS AND METHODS

2.1. BIOLOGICAL EXPERIMENTS

Biological experiments were described in details elsewhere [17, 18]. Briefly, the pIF 105 plasmid was treated with 100 and 200 mM CAA for ethenoadducts formation (the highest 500 mM CAA was not analyzed in this paper). Modified and mock treated plasmids were introduced into bacterial cells prepared in two versions: (i) non-adapted or (ii) adapted with methyl methanesulphonate (MMS, $CH_3SO_3CH_3$, as an alkylating agent) for inducing the adaptive response system. Each experiment was carried out three times and mutagenesis level was defined as mutation frequencies (MF). The MF was calculated as the number of mutants per 10^4 transformed cells.

2.2. DATA AND PERFORMED ANALYSIS

The mutagenesis levels were determined in some *E.coli* strains: (i) wild-type (*wt*), (ii) alkA-deficient (*alkA*), and (iii) alkB-deficient (*alkB*). The strains were studied in following conditions: (i) non- and adapted with MMS (denoted as *NA* and *A* condition, respectively), and (ii) no DNA modification and after DNA modification (mock or CAA treated plasmids), independently for *NA* and *A* condition. In experiments 7 to 28 values of mutagenesis level were obtained. Statistical analysis was done with the use of Mann-Whitney test for comparison the *E.coli* strains (Statistica v.6, StatSoft). Data were present as medians with interquartile ranges. The fuzzy *k*-NN classifier [28] was used to differentiate the strains in studied experimental situations. Lower misclassification rate (error rate, E_r) corresponds to better distinguishing [9, 12]. The probability of misclassification was estimated by the *leave one out* method [7]. In our studies the ROC (i.e. Receiver-Operating Characteristic) analysis was also included [29].

3. RESULTS

Table 1 presents medians with interquartile ranges obtained from the molecular experiments. Table 2 gives error rates (E_r) of differentiation between the strains without- (only spontaneous mutations) and Table 3 with CAA treatment, when εA is induced in non-adapted (*NA)* and adapted (*A)* condition. These results are obtained by the use of fuzzy *k*-NN classifier. Figures 1-3 and Table 4 show results of ROC analysis.

Table 1. Medians with interquartile range of mutation frequency (MF, number of mutants per 10^4 transformed cells) in the *E.coli* strains.

STRAINS:	[*NA*] CONDITION	[*A*] CONDITION
wt:		
no DNA modification	1.5 (0.4-2.2)	0.6 (0.4-0.8)
100 mM CAA treatment	12.7 (10.1-18.8) **	2.0 (1.7-10.1)
200 mM CAA treatment	20.2 (16.7-26.5) **	5.7 (2.3-12.5)
alkA:		
no DNA modification	1.1 (1.0-1.7) *	1.8 (1.4-2.2)
100 mM CAA treatment	13.9 (12.0-17.5) **	7.0 (1.3-11.4)
200 mM CAA treatment	19.4 (13.6-22.6) **	7.3 (4.5-15.2)
alkB:		
no DNA modification	8.0 (4.8-15.5) #, +	8.3 (7.3-12.1) #, +
100 mM CAA treatment	48.6 (25.0-95.2) #, +	33.3 (20.0-60.0) #, +
200 mM CAA treatment	52.6 (52.6-127.7) #, +	104.0 (52.0-160) #, +

* $p<0.05$, ** $p<0.005$ for [*NA*] *vs.* [*A*] under the same experimental procedure
# $p<0.005$ for wild type *vs*. other strains under the same experimental procedure
+ $p<0.005$ *alkA vs. alkB* under the same experimental procedure

Statistically significant differences between *NA* and *A* conditions were observed for wild-type strain after DNA modification and for *alkA* strain without and with DNA modification. Whereas these difference of *alkB* strains were not statistically significant (Table 1). On the other hand, *alkB* meaningfully differed from other strains ($p<0.005$) and its mutagenesis level always achieved the highest value.

Table 2. Misclassification rates (E_r). Differentiation between strains without DNA modification, separately for *NA* and *A* conditions, by the fuzzy *k*-NN classifier.

DIFFERENTIATION OF STRAINS:	[*NA*] CONDITION	[*A*] CONDITION
wild type vs.:		
alkA	0.161	0.000
alkB	0.029	0.079
Between strains:		
(alkA,alkB)	0.000	0.055

Table 3. Misclassification rates (E_r). Differentiation between strains after DNA modification by CAA, separately *NA* and *A* conditions, by the fuzzy *k*-NN classifier.

DIFFERENTIATION OF STRAINS:	CAA [*NA*]		CAA [*A*]	
	100mM	200Mm	100mM	200Mm
wt-type vs.:				
alkA	0.250	0.353	0.263	0.162
alkB	0.071	0.000	0.000	0.000
Between strains:				
(alkA,alkB)	0.036	0.000	0.063	0.000

The obtained results of *k*-NN analysis are in accordance with the above mentioned statement that *alkB* strain differed the most from others (Tables 2 and 3). After the DNA modification by 200 mM CAA (Table 3), *alkB* can be correctly recognized as for *NA* as well for *A* (E_r=0.000). So, it seems that the CAA concentration is optimal to study DNA repair in proposed *E.coli* system.

The ROC curves were made for the strains without and with DNA modification by 200mM CAA (Fig. 1-3). Sensitivity, specificity and accuracy of the mutagenesis test system are presented in Table 4. The wild-type strain was considered as the control strain.

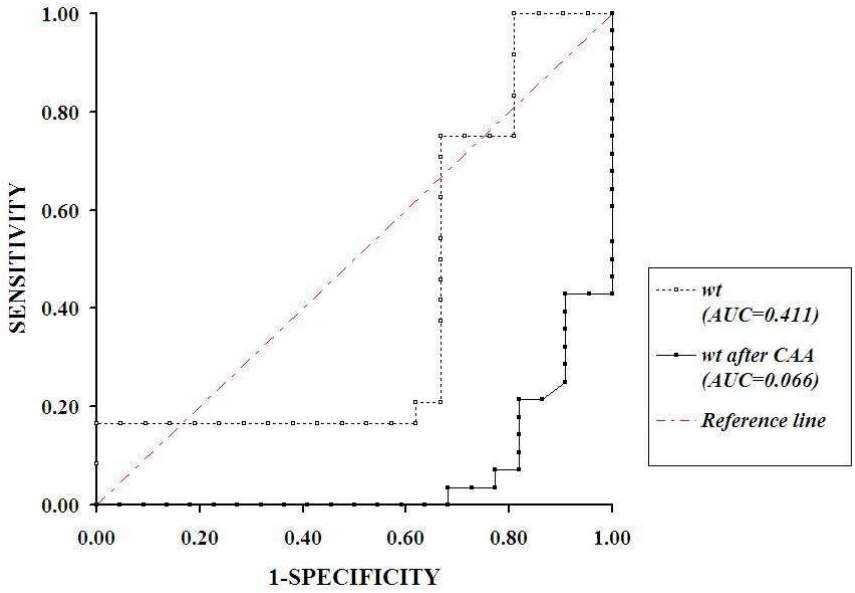

Fig. 1. The ROC curves for wild-type strain without (dashed line) and with DNA modification by 200 mM CAA (continuous line) for adapted *vs*. non-adapted condition.

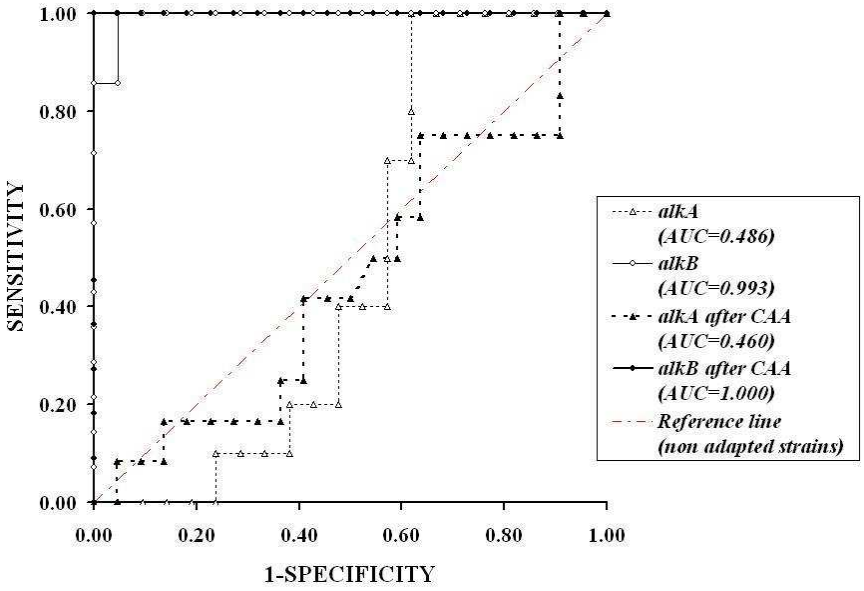

Fig. 2. The ROC curves for *alkA* and *alkB* without (dashed line) and with DNA modification by 200 mM CAA (continuous line) in non-adapted condition.

AUC (i.e. Area Under the Curve) of wild-type achieved only value of 0.411 and it dramatically decreased to 0.006 after DNA modification (Fig. 1), as it was expected for spontaneously mutations.

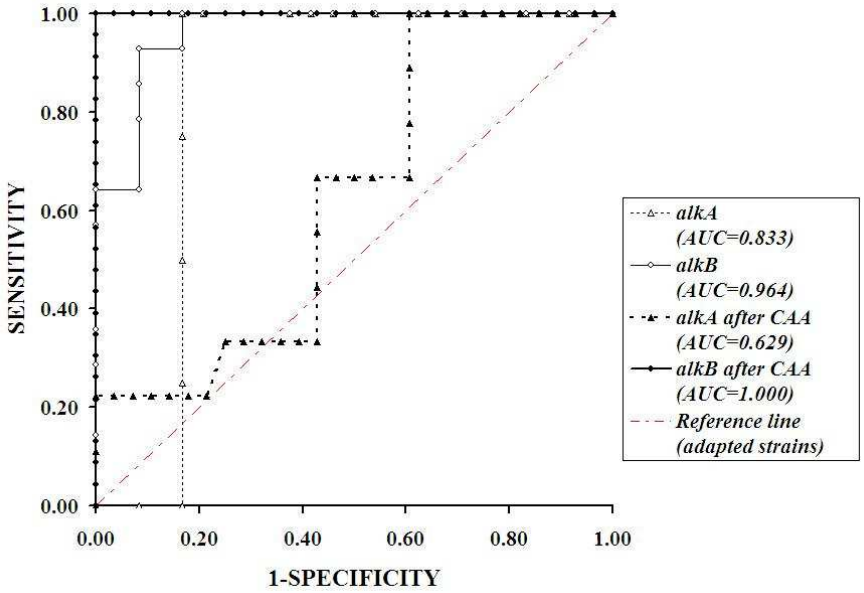

Fig. 3. The ROC curves for *alkA* and *alkB* without (dashed line) and with DNA modification by 200 mM CAA (continuous line) in adapted condition.

AUC of *alkA* was below 0.500 without and with DNA modification in non-adapted condition (Fig. 2). After adaptation, these values were 0.833 and 0.625, respectively (Fig. 3). Independently on the adaptation, AUC of *alkB* achieved values above 0.900 for no DNA modification and 1.00 after CAA treatment (Fig. 2 and 3). These results point out that *alkB* mutagenesis level is the best marker for the mutagenesis test system.

Table 4. Sensitivity, specificity and accuracy for optimal cut-off values in mutagenesis tests of *E.coli* model.

STRAINS	SENSITIVITY	SPECIFICITY	ACCURACY	CUT-OFF VALUE
NON-ADAPTED CONDITION				
AlkB				
no DNA modification	1.00	0.95	0.97	3.4
DNA modification	1.00	1.00	1.00	46.2
AlkA				
no DNA modification	1.00	0.38	0.58	0.9
DNA modification	1.00	0.09	0.41	10.3
ADAPTED CONDITION				
AlkB				
no DNA modification	1.00	0.83	0.89	2.1
DNA modification	1.00	1.00	1.00	24.2
AlkA				
no DNA modification	1.00	0.83	0.88	1.1
DNA modification	1.00	0.39	0.54	3.75

The performed analysis indicated that AlkB protein is strongly engaged in εA repair processes in the *E.coli* cells. After DNA modification we have observed: i) the highest values of *alkB* mutation level, ii) statistically significant differences of *alkB vs.* other strains ($p<0.005$), iii) quiet correct recognition of *alkB vs.* wild-type and *alkA* strains after 200 mM CAA treatment ($E_r=0.000$), iv) sensitivity, specificity and accuracy equal to 1.00.

4. DISCUSSION AND CONCLUSION

In the present paper we used pIF105 plasmid, which is indicatory for AT→TA substitutions caused by εA and we studied participation of *alkB* and *alkA* gene products in repair of 1,N^6-ethenoadenine in *E.coli* cells. Although *in vitro* data indicated that εA is a very poor substrate of AlkA glycosylase, we included AlkA to our study in order to estimate its possible role *in vivo*. Therefore, we considered the wild-type *E.coli* and *alkB*-, and *alkA*-deficient strains. Since AlkA and AlkB are inducible with Ada response proteins, we analyzed mutagenesis in non-adapted and adapted (with a low dose of MMS) *E.coli* cells. The performed analysis demonstrated mainly that:

- After DNA modification, the *alkB* strain was well recognized, especially in experiments with induction of the adaptive response, and its mutation level increased the most.
- The ROC analysis clearly showed that *alkB* is actively included in the repair processes.

Resuming, the obtained results confirmed significant participation of AlkB protein in εA repair. In addition, it seems that AlkA protein is weakly involved in the repair. Thus, εA is repaired by AlkB *in vivo* and AlkA have small significance in the εA repair system. It has been also corroborated in our observations for other mutants of *E.coli* strains [18]. Presented application of the mutagenesis test system is effective and helpful method in the bacterial model study of DNA repair mechanisms.

5. ACKNOWLEDGMENTS

We would like to thank Drs. Iwona Fijałkowska and Piotr Jonczyk from our Institute for kindly providing the pIF105 plasmid for mutagenesis tests. This work was supported by grant N301 065 31/1979 of the Ministry of Science and Higher Education, and partly by the Polish-Norwegian grant PNRF-143-AI-1/07.

BIBLIOGRAPHY

[1] ALTIERI F., GRILLO C., MACERONI M., CHICHIARELLI S., DNA damage and repair: from molecular mechanisms to health implications, Antioxidants & Redox Signaling 2008, Vol. 10, No. 5, pp. 1-47.

[2] BARNES D.E., LINDHAL T., Repair and genetic consequences of endogenous DNA base damage in mammalian cells, Annual Review Genetic 2004, 38, pp. 445-476.

[3] BARTSCH H., Keynote address: Exocyclic adducts as new risk markers for DNA damage in man. In: Exocyclic DNA Adducts in Mutagenesis and Carcinogenesis, SINGER B., BARTSCH H, (eds.), IARC Scientific Publication, No. 150, Lyon, 1999, pp. 1-16.

[4] BOLT HM., Vinyl chloride - a classical industrial toxicant a new interest, Critical Reviews in Toxicology 2005, 35, pp. 307-323.
[5] CUPPLES C.G., MILLER JH., A set of lacZ mutations in Escherichia coli that allowed rapid detection of each the six base substitutions, Proc Natl. Sci. USA 1989, Vol. 86, pp. 5345-5349.
[6] DELANEY JC., SMEESTER L., WONG C., FRICK LE., TAGHIZADEH K., WISHNOK JS., DRENNAN CL., SAMSON L.D., ESSIGMANN JM., AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo, Nature Structural & Molelular Biolology 2005, Vol. 12, pp. 855-860.
[7] DEVIJVER P.A., KITTLER J., Pattern Recognition: A Statistical Approach, Prentice Hall, London, 1982.
[8] DRABLOS F., FEYZI E., AAS PA., VAAGBO C.B., KAVLI B., BRATLIE M.S., PENA-DIAZ J., OTTERLEI M., SLUPPHAUNG G., KROKAN H.E., Alkylation damage in DNA and RNA-repair mechanisms and medical significance, DNA Repair 2004, Vol. 3, pp. 1389-1407.
[9] DUDA R.O., HART PE., STOCK D.G., Pattern Classification, John Wiley and Sons, New York, 2001.
[10] EVANS M.D., DIZDAROGLU M., COOKE M.S., Oxidative DNA damage and disease: induction, repair and significance, Mutation Research 2004, 567, pp. 1-61.
[11] FIJALKOWSKA, I.J., JONCZYK, P., TKACZYK, M.M., BIALOSKORSKA, M., SCHAAPER, R.M., Unequal fidelity of leading strand and lagging strand DNA replication on the Escherichia coli chromosome, Proc. Natl. Acad. Sci. USA 1998, Vol. 95, pp. 10020-10025.
[12] FIX E., HODGES J.L., Discriminatory Analysis. Nonparametric Discrimination Small Sample Performance, project 21-49-004, Report Number 11, USAF School of Aviation Medicine, Randolph Field, Texas 1952, 280-322, reprinted in the book: Dasarathy B.V., NN Pattern Classification Techniques, IEEE Computer Society Press, 1991, pp. 40-56.
[13] HUGHES SJ., SEDGWICK B., The adaptive response to alkylation damage, The Journal of Biological Chemistry, Vol. 264, No. 35, 1989, pp. 21369-21375.
[14] IRIGARAY P., BELPOMME D., Basic properties and molecular mechanisms of exogenous chemical carcinogens, Carcinogenesis 2010, Vol. 31, No. 2, pp. 135-148.
[15] JACOBSEN JS., PERKINS CP., CALLAHAN JT., SAMBAMURITI K., HUMAYUN MZ., Mechanisms of mutagenesis by chloroacetaldehyde, Genetics 1989, Vol. 121, pp. 213-232.
[16] MACIEJEWSKA A.M., JÓŹWIK A., KUŚMIEREK J.T., SOKOŁOWSKA B., Application of the k-NN classifier for mutagenesis tests. Recognition of wild type and defective in DNA repair bacterial strains on the basis of adaptive response to alkylating agents, Biocybernetics and Biomedical Engineering, Vol. 28, No. 3, 2008, pp. 45-50.
[17] MACIEJEWSKA A.M., RUSZEL K.P., NIEMINUSZCZY J., LEWICKA J., SOKOŁOWSKA B., GRZESIUK E., KUŚMIEREK J.T., Chloroacetaldehyde-induced mutagenesis in Escherichia coli: the role of AlkB protein in repair of 3,N4-ethenocytosine and 3,N4-α-hydroxyethanocytosine, Mutation Research 2010, 684, pp. 24-34.
[18] MACIEJEWSKA A.M., SOKOŁOWSKA B., NOWICKI A., KUŚMIEREK J.T., The role of AlkB protein in repair of 1,N6-ethenoadenine in Escherichia coli cells, Mutagenesis 2011, Vol. 26, No. 3, pp. 401-406.
[19] MARTIN G.M., AUSTAD S.N., JOHNSON T.E., Genetic analysis of ageing: role of oxidative damage and environmental stress, Nature Genetic 1996, Vol. 13, pp. 25-34.
[20] MISHINA Y, YANG C.G, HE C., Direct repair of the exocyclic DNA adduct 1,N6-ethenoadenine by the DNA repair AlkB proteins, Journal of the American Chemical Society, Vol. 127, 1989, pp. 14594-14595.
[21] MROCZKOWSKA M.M., KOLASA I.K., KUSMIEREK J.T., Chloroacetaldehyde-induced mutagenesis in Escherichia coli: specificity of mutations and modulation by induction of the adaptive response to alkylating agents, Mutagenesis 1993, Vol. 8, pp. 341-348.
[22] NAIR J., BARBIN A., VELIC I., BARTSCH H., Etheno DNA-base adducts from endogenous reactive species, Mutation Research 1999, Vol. 424, pp. 59-69.
[23] NIEMINUSZCZY J., GRZESIUK E., Bacterial DNA repair genes and their eukaryotic homologues: 3. AlkB dioxygenase and Ada methyltransferase in the direct repair of alkylated DNA, Acta Biochimica 2007, Vol. 54, No. 3, pp. 459-468.
[24] ROLING R.L., MCKINNON P.J., Linking DNA damage and neurodegeneration. Trends in Neuroscience 2000, Vol. 23, No. 9, pp. 417-424.
[25] SEDGWICK B., BATES P.A., PAIK J., JACOBS S.C., LINDAHL T., Repair of alkylated DNA: recent advances, DNA Repair 2007, Vol. 6, pp. 429-442.
[26] SEDGWICK B., LINDAHL T., Recent progress on the Ada response for inducible repair of DNA alkylation damage, Oncogene 2002, Vol. 21, pp. 8886-8894.
[27] SHRIVASTAV N., LI D., M.ESSIGMANN J., Chemical biology of mutagenesis of DNA repair: cellular responses to DNA alkylation, Carcinogenesis 2010, Vol. 31, No. 1, pp. 59-70.
[28] SOKOŁOWSKA B., JÓŹWIK A., POKORSKI M., A fuzzy-classifier system to distinguish respiratory patterns evolving after diaphragm paralysis in the cat, Japanese Journal of Physiology 2003, Vol. 53, No. 4, pp. 301-307.
[29] ZWEIG M.H., CAMPBELL G., Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine, Clinical Chemistry 1993, Vol. 39, No. 4, pp. 561-577.