Transactions on Ecology and the Environment vol 46, © 2001 WIT Press, www.witpress.com, ISSN 1743-3541

Ecotoxicology modeling of methylmercury bioaccumulation

G. K. Luk
Department of Civil Engineering, Ryerson University, Canada

Abstract

Inorganic elemental mercury discharged into the natural aquatic system is converted under suitable environmental conditions into a highly toxic, organic form of mercury called methylmercury. Studies have found that this form of mercury is more readily available to biota, and may be efficiently transferred up the food chain, thereby concentrating in larger, predatorial fish species in the top trophic levels. The objective of this paper is to apply a bioenergetics-based mathematical model to describe the bioaccumulation behavior of methylmercury in fish. Model results are compared to field observation collected from various locations of Lake Ontario, and the general body buildup of mercury with increasing age is explained through the various pathways of transfer in conjunction with the bioenergetics concepts. The study is conducted on two contrasting Lake Ontario salmonid species: Walleye (*Stizostedion vitreum vitreum*) and Yellow Perch (*Perca flavescens*), occupying different trophic positions in the aquatic food chain. In spite of the notable difference in the fish size, metabolism, diet and trophic level, it was demonstrated that methylmercury accumulation for both species is described satisfactorily by the model.

1 Introduction

Among the various pollutants pervasive in the natural aquatic system, mercury has always received special attention due to its toxic properties. In the presence of methylating bacteria and under suitable environmental conditions, inorganic elemental mercury is converted into a highly toxic, organic form of mercury called methylmercury. Studies have found that this form of mercury is more readily available to biota, and may be efficiently transferred up the food chain, thereby concentrating in larger, predatorial fish species in the top trophic levels. This process of bioaccumulation explains the fact that even though the level of

toxins in the environment is relatively low, the concentrations found in the fish are typically many times higher.

Both inorganic and methylmercury are taken up directly from water, sediments and food. The uptake and clearance of mercury depend on the form and source of mercury, as well as the type of receptor tissue, resulting in different patterns of accumulation. Methylmercury is more efficiently accumulated than inorganic mercury for most aquatic organisms [1]. It is readily transferred across biological membranes. Within the organism, methylmercury is strongly bound to sulfhydryl groups in proteins of tissues such as muscle, and is much slower to clearance than inorganic mercury. Thus, methylmercury has a much greater potential for bioaccumulation and a longer half-life in organisms than inorganic mercury.

The accumulation of mercury from water occurs via the gill membranes. Gills take up aqueous methylmercury more readily than inorganic mercury, from which it is eventually transferred to the muscles and other tissues and retained for long periods of time [2]. In contrast, inorganic mercury taken up with food initially accumulates in the tissues of the posterior intestine of fish, from which only limited transfer to other parts of the body takes place. Studies [3] show that 80% of accumulated elemental mercury would depurate from the fish intestine in as short as 15 days. As a result, the liver and kidney in fish tend to have higher percentages of inorganic mercury than muscle tissue, although percentages vary by organ and species [2].

2 Objectives

To describe the body burden of methylmercury in fish, a bioenergetics-based model developed by Luk [4] is used. The program, written in Visual Basic 4.0, is specifically designed for the simulation of body concentrations of methylmercury. Two kinds of sport fish commonly found in Lake Ontario are selected. The first one, Walleye (*Stizostedion vitreum vitreum*), is probably the most economically valuable fish species in Canada's inland waters. It is a major commercial sport fish in the provinces of Ontario, Quebec and the Prairies. Walleye occupies a high position in the aquatic food chain and is predacious. It is characterized as a relatively large fish, spanning in sizes from 25-85 cm long and weighs up to 6 kg. The second species is Yellow Perch (*Perca flavescens*), which has long been considered as prime importance to man. Yellow Perch inhabits a vast territory and congregates near shore in the spring. All of these factors make it readily available to fishermen, both commercial and recreational. Yellow Perch belongs to a much lower trophic level in the aquatic ecosystem, with lengths from 15-30 cm and weights from 0.3-0.6 kg, and is primarily planktivorous.

The database of this study is obtained from the "Sport Fish Contaminant Monitoring Program", operated by the Ontario Ministry of Environment and Energy. The data set, comprised of 419 Walleye and 686 Yellow Perch samples, was collected for a ten-year period from 1989 to 1998 at various points in Lake Ontario. It contains information on the length, weight, sex and lipid fraction of

the fish, the date of collection, the location of the sampling site, and the concentration of methylmercury in the dorsal filet.

3 The bioaccumulation model

3.1 Bioenergetics equation

Pollutant uptake comes from food and water, the amounts of which are related to the energy requirements of the fish. The total amount of energy required by the fish, commonly referred to energy for total metabolism, is given by the equation:

$$Q = Q_{lr} + Q_c, \tag{1}$$

where Q_{lr} = the "low-routine metabolism" energy cost, incorporating all the activity-related energy such as swimming and foraging, and Q_c = the energy cost for utilization of food (e.g. digestion and absorption). It was demonstrated [5] that the low-routine metabolism may be represented empirically as a power-function of the weight of the fish as:

$$Q_{lr} = \alpha_{lr} W^{\tau}, \tag{2}$$

where α_{lr} = the low-routine coefficient, W = the weight of the fish, and τ = the body-weight exponent for metabolism.

The energy cost for utilization of food (Q_c) is directly proportional to the rate of growth of fish, (dW/dt). It implies that regardless of the actual fish size, a fish with a steady growth rate will have a lower Q_c than an actively growing fish. Taking energy equivalent of the flesh of fish into account, the equation for Q_c may be represented by:

$$Q_c = q_f \beta \left(\frac{dW}{dt}\right), \tag{3}$$

where β = the proportionality constant, and q_f = the energy equivalent (energy content) of the flesh of fish. The final form of the bioenergetics equation is obtained by substituting eqns (2) and (3) into eqn (1) to get:

$$Q = \alpha_{lr} W^{\tau} + q_f \beta \left(\frac{dW}{dt}\right). \tag{4}$$

3.2 Uptake of Pollutants from food

The amount of energy a fish obtained from its food (ration) can be expressed as:

$$Q_{food} = q_{fd} E_{fd} R, \tag{5}$$

where q_{fd} = the energy equivalence of the prey (food), E_{fd} = the efficiency of assimilation of food, and R = the ration consumed by the fish. The efficiency of food assimilation is important because some of the food energy is wasted, egested or excreted, and is therefore not available for use. This amount of food energy (Q_{food}) is used to support all the required metabolic activities (total metabolism) as well as growth of the fish, or:

$$Q_{food} = Q + q_f \left(\frac{dW}{dt}\right). \tag{6}$$

Equating eqns (5) and (6) and solving for Q, one obtains:

$$Q = q_{fd}\, E_{fd}\, R - q_f\left(\frac{dW}{dt}\right). \tag{7}$$

Both eqns (4) and (7) represent the total metabolic cost of a fish, the former from energy balance and the latter from ration requirement. Equating the two equations, the following expression for ration is obtained:

$$R = \frac{1}{q_{fd}\, E_{fd}}\left[\alpha_{lr}\, W^{\tau} + q_f\,(\beta+1)\frac{dW}{dt}\right]. \tag{8}$$

Toxicants such as mercury are consumed with contaminated food and absorbed from the gastrointestinal tract into the body. They are then deposited and stored in various tissues particularly in muscles. Mercury will enter into the body of fish either directly from solution or from food in the form of CH_3Hg^+ [6]. With prolonged exposure to contaminated environment and continuous intake of contaminated food, the tissue concentration of methylmercury will increase as a fish grows. For this reason, not only the amounts but the type and composition of a fish's diet are important. Since methylmercury is lipophilic in nature, it binds with the sulfhydryl groups of proteins strongly. Therefore, a diet with a high lipid content will result in a higher concentration of methylmercury. Fishes like Walleye and Lake Trout, that occupy a high trophic level in the food chain, tend to consume more contaminated prey and therefore the accumulation is increased. The rate of pollutant uptake through the food pathway can be expressed as:

$$\left(\frac{dP}{dt}\right)_f = E_{pf}\, R\, C_{pf}\,, \tag{9}$$

where P = the total body burden of pollutant, E_{pf} = the efficiency of pollutant uptake from food, R = the ration consumed by the fish, and C_{pf} = the concentration of pollutant in food. Replacing the expression of ration from eqn (8) into (9) yields:

$$\left(\frac{dP}{dt}\right)_f = \frac{E_{pf}\, C_{pf}}{q_{fd}\, E_{fd}}\left[\alpha_{lr}\, W^{\tau} + q_f\,(\beta+1)\frac{dW}{dt}\right]. \tag{10}$$

3.3 Uptake of Pollutants from Water

The process of diffusion, as described by the Fick's law, governs the absorption of oxygen and release of carbon dioxide from the fish's body during respiration. In general, the gas diffuses into an aqueous layer covering the epithelial cells lining the respiratory system of the organism. The diffusion is driven by the concentration gradient between the interior of the fish and its surroundings. The amount of energy that is obtained from the intake of oxygen is expressed as:

$$Q = E_{ox}\, C_{ox}\, q_{ox}\, V, \tag{11}$$

where E_{ox} = the absorption efficiency of oxygen, C_{ox} = the concentration of dissolved oxygen in water, q_{ox} = the energy equivalence of oxygen, and V = the volume of water passing through the gills. With this equation, the rate of pollutant uptake through the water pathway may be expressed as:

$$\left(\frac{dP}{dt}\right)_w = E_{pw}\, C_{pw}\, V = \frac{E_{pw}\, C_{pw}\, Q}{E_{ox}\, C_{ox}\, q_{ox}}, \tag{12}$$

where E_{pw} = the efficiency of transfer of pollutants from water, and C_{pw} = the concentration of pollutants in water. By substituting the expression of Q from eqn (4) into eqn (12), the rate of pollutant uptake through the water pathway is obtained as:

$$\left(\frac{dP}{dt}\right)_w = \frac{E_{pw}\, C_{pw}}{E_{ox}\, C_{ox}\, q_{ox}}\left(\alpha_{lr}\, W^{\tau} + q_f \beta \frac{dW}{dt}\right). \tag{13}$$

3.4 Clearance

Previous studies on the excretion of methylmercury from fish demonstrated that it is a slow process, with half-lives ranging from one to three or more years [7]. The whole body clearance of methylmercury has been shown [8] to follow first-order kinetics under various sets of environmental and physiological conditions, with very little dependence on temperature and metabolic rate. A general expression for clearance is given as:

$$\left(\frac{dP}{dt}\right)_{cl} = -k\,P, \tag{14}$$

where k = the clearance coefficient from excretion and growth dilution.

The overall rate of change of pollutant burden in the body of fish is governed by the pollutant intake from the food pathway, the pollutant intake from the water (or dissolved oxygen) pathway, and the clearance. Combining eqns (10), (13) and (14), the following bioaccumulation equation is obtained:

$$\left(\frac{dP}{dt}\right) = \frac{E_{pf}\, C_{pf}}{q_{fd}\, E_{fd}}\left[\alpha_{lr}\, W^{\tau} + q_f(\beta+1)\frac{dW}{dt}\right] + \frac{E_{pw}\, C_{pw}}{E_{ox}\, C_{ox}\, q_{ox}}\left(\alpha_{lr}\, W^{\tau} + q_f \beta \frac{dW}{dt}\right) - k\,P\,. \tag{15}$$

4 Concentration of mercury in food

Mercury concentration in food is dependent on the diet of the fish; those that are predacious and feed on other small fishes tend to have a higher C_{pf} than those that feed exclusively on zooplankton and invertebrates. Another factor to be considered is that changes in diet pattern usually occur as a fish grows. In the case of Walleye, for example, the major diet items switch from mayfly nymph and crayfish to Rainbow Smelt and Yellow Perch by the age-group of 3. Table 1 shows a reconstruction of Walleye's diet by Mathers and Johansen [9], and the change this has on the overall methylmercury concentration from food. Yellow Perch, on the other hand, has a much simpler diet pattern, with its major food item as zooplankton (*Daphnia pulex*) [10]. The proposed C_{pf} value for this is 0.03 μg/g after Snodgrass and Harris [11].

Table 1: The diet composition of Walleye (after Mathers & Johansen [9])

	Diet items	% by weight	Mean q_{fd} (kcal/g)	C_{pf} (ug/g)
Juvenile age(0-3)	Mayfly Nymph	50	1.0	0.047
	Crayfish	50	1.0	0.042
Adult Age > 3	Rainbow Smelt	69	1.36	0.253 (0.07 - 0.570)
	Yellow Perch	19	1.36	0.107 (0.04 - 0.230)
	Emerald Shiner	9	1.36	0.107 (0.120 - 0.150)
	Common Shiner	3	1.36	0.043 (0.030 - 0.060)

5 Model results

A summary of the parameters of the model for the two fish species is provided in Table 2. The field data, as obtained from the Sport Fish Contaminant Monitoring Program, represents the methylmercury concentration of a skinned filet taken from the dorsal tissue of the fish. However, the model simulates the whole-fish methylmercury concentration, which is generally higher because of the lipids it contains. Therefore, adjustment should be made to the field data to convert them from filet to whole fish concentrations. Amrhein et al. [12] did a comparison study on the average whole-fish to filet PCB concentration ratios, and found it to be 1.70 for Coho Salmon and 1.47 for Rainbow Trout. In the lack of any better information, and based on the similar lipid soluble nature of between methylmercury and PCBs, a conversion factor of 1.50 is adopted for this study.

The results from the model, together with the converted field data (discrete data points and 90% confidence interval), are shown in Figs. 1 and 2. It is observed from the figures that the bioaccumulation model is quite effective in predicting the trend of mercury uptakes in both species. Despite the vast difference of genetic makeup and trophic levels of the species, the model has successfully produced predictions consistent with the field data. The following observations are made from the model results:

1. Methylmercury concentration in the fish body increases with the fish age and size. This is reflected from the model and observed field data. Considering there is an effect of growth dilution, which theoretically reduces the concentration with weight increases, this increase is quite remarkable. The rate of accumulation must have increased very rapidly during the course of growth, so much so that the growth-dilution effect is completely masked.
2. It is observed from field data that the maximum methylmercury concentration in Walleye is five times that of Yellow Perch. The model is able to reproduce this trend perfectly. Since both species are sampled from Lake Ontario, the fishes are exposed to a similar aquatic environment. As a result, the substantial difference in their body concentrations is derived exclusively from the difference in metabolism and diet composition. Therefore, it may be implied that the overall bioaccumulation of

methylmercury for a species is very much dependent on the genetic makeup (affecting metabolism) and trophic levels (affecting diet preference).

3. Figure 1 shows an abrupt increase of methylmercury in Walleye at the age of three. This is really a reflection of the diet reconstruction, and therefore more of a modeling artifact than realistic representation. It is true that Walleye changes its diet during the course of growth, but this change is more likely to occur gradually. For planktivorous fish like Yellow Perch, diet is quite consistent over their life span. As a result, the model result produces a smooth curve of methylmercury accumulation. In light of this, more efforts should be devoted to produce better and more realistic reconstruction of the diet of individual species, in order to improve the overall predictive ability of the model.

Table 2: Summary of model parameters

Symbol	Unit	Parameter value	
		Walleye	Yellow Perch
Fish Growth			
$W\infty$	g	5967	550
k	wk^{-1}	0.00453	0.00315
b	----	3.3196	3.2206
Clearance			
k	wk^{-1}	0.002	0.002
Metabolic Parameters			
$\propto_{lr}$	kcal/(wk.g^{τ})	0.238	0.125
τ		0.77	0.81
Energy Equivalence			
q_f	kcal/g	1.145	1.250
q_{fd}	kcal/g	Table 1	1.0
q_{ox}	kcal/ g O_2	3.42	3.42
Efficiency Factors			
E_{fd}	----	0.75	0.675
E_{ox}	----	0.75	0.75
E_{pf}	----	0.7	0.80
E_{pw}	----	0.12	0.12
Concentrations			
C_{pf}	μg/g	Table 1	0.03
C_{pw}	ppb	0.05	0.05
C_{ox}	μg/g	10.9	10.9

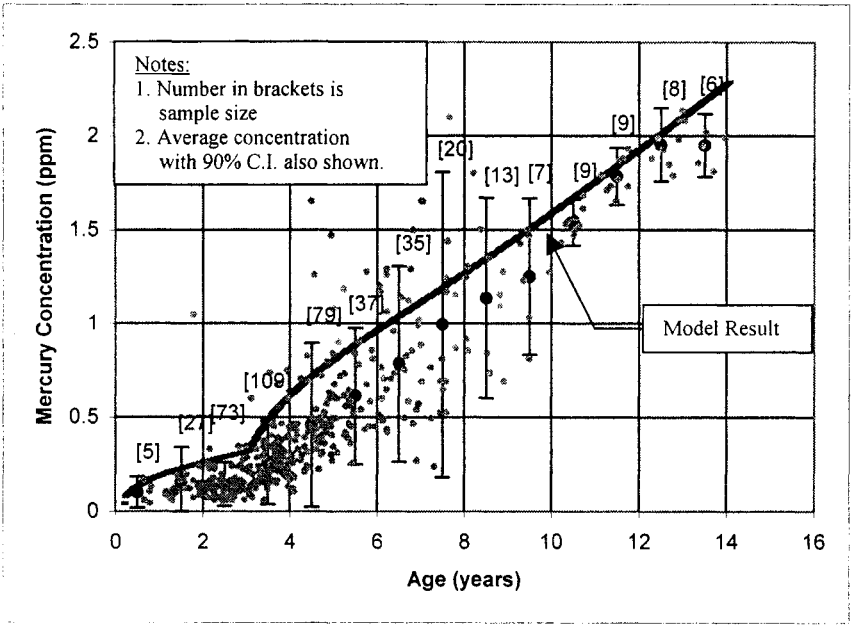

Figure 1: Model results and field data for Walleye.

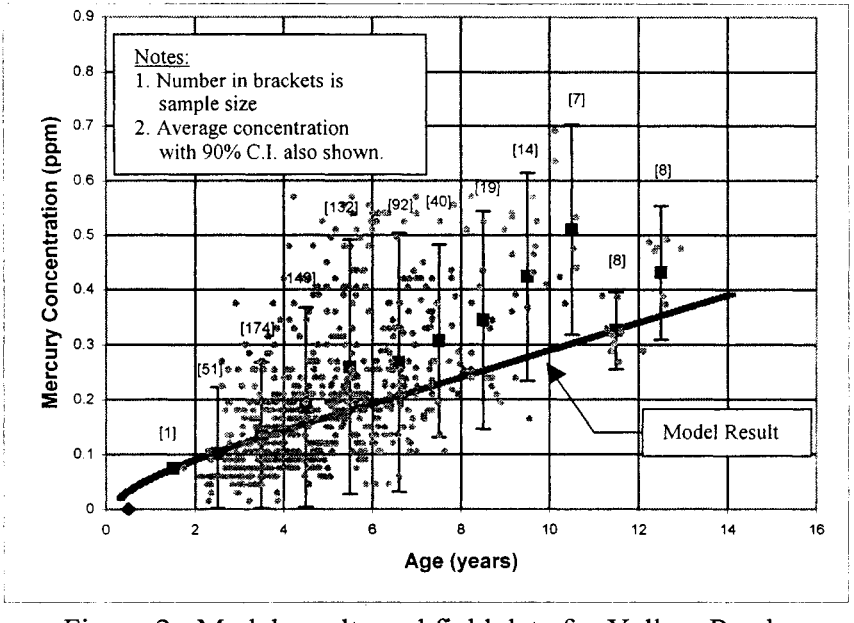

Figure 2: Model results and field data for Yellow Perch.

6 Conclusions

The bioenergetics-based bioaccumulation model provides a useful framework to study the accumulation of methylmercury in fish, with the potential to mechanistically accommodate system-to-system variations among species and habitats. Results from the model suggest that food is clearly the dominant mercury uptake pathway for both Walleye and Yellow Perch. The variation in the total body burden levels among different species may be explained by their biological characteristics as described by the growth pattern, activity level, metabolism and diet preference. In general, the higher is the position of a fish in the aquatic food chain, the greater is its likelihood to exhibit a high level of mercury in the body.

Acknowledgement

This study is supported by the Natural Science and Research Council (NSERC) of Canada. Technical assistance provided by the Research Assistant Mr. Kenny Ling is greatly appreciated.

References

[1] Mason, R.P., Reinfelder, J.R. & Morel, F.M.M. Bioaccumulation of mercury and methylmercury. *Water, Air & Soil Pollution,* **80**, pp. 915-921, 1995.

[2] Hansen, C.T., Nielsen, C.O., Dietz, R. & Hansen, M.M. Zinc, cadmium, mercury and selenium in Minke Whales, Belugas and Narwhals from West Greenland. *Polar Biology*, **10**, pp. 529-539, 1990.

[3] Boudou, A., Delnomdedieu, M., Georgescauld, D., Ribeyre, F. & Saouter, E. Fundamental roles of biological barriers in mercury accumulation and transfer in freshwater ecosystems (analysis at organism, organ, cell and molecular levels). *Water, Air & Soil Pollution,* **56**, pp. 807-822, 1991.

[4] Luk, G.K. & Brockway, F. Application of a PCBs bioaccumulation model to Lake Ontario Lake Trout. *J. of Ecological Modelling*, **101**, pp. 97-111, 1997.

[5] Borgmann, U. & Whittle, D.M. Bioenergetics and PCB, DDE, and mercury dynamics in Lake Ontario Lake Trout. *Can. J. Fish. Aquat. Sci.,* **49**, pp. 1086-1096, 1992.

[6] Jackson, T.A. Biological and environmental control of mercury accumulation by fish in lakes and reservoirs of Northern Manitoba, Canada. *Can. J. Fish. Aquat. Sci.,* **48**, pp. 2449-2470, 1991.

[7] McKim, J.M., Olson, G.F., Holcombe, G.W. & Hunt, E.P. Long-term effects of methylmercuric chloride on three generations of brook trout *(Salvelinus fontinalis*): toxicity, accumulation, distribution, and elimination. *J. Fish. Res. Board of Canada,* **33**, pp. 2726-2739, 1976.

[8] De Freitas, A.S.W., Qadri, S.U. & Chase, B.E. Origins and fate of mercury compounds in fish. *Proc. Int. Conf. On Transport of Persistent Chemicals in Aquatic Ecosystems*, Ottawa, Canada, pp. 31-36, 1974.

[9] Mathers, R. A. & Johansen, P.H. The effects of feeding ecology on mercury accumulation in Walleye (*Stizostedion vitreum*) and Pike (*Esox lucius*) in Lake Simcoe. *Can. J. Zool.*, **63**, pp. 2006-2012, 1985.

[10] Mills, E.L. & Forney, J.L. Energetics, food consumption, and growth of young Yellow Perch in Oneida Lake, New York. *Trans. of Am.Fish. Soc.*, **110**, pp. 479-488, 1981.

[11] Snodgrass, W.J. & Harris, R.C. Bioenergetics simulation of mercury uptake and retention in Walleye (*Stizostedion vitreum*) and Yellow Perch *(Perca flavescens*). *Water Poll. Res. J. Canada,* **28(1)**, pp. 217-236, 1993.

[12] Amrhein, J.F., Stow, C.A. & Wible, C. Whole-fish versus filet PCB concentrations: an analysis using classification and regression trèe models. *Env. Toxicology & Chemistry,* **18(8)**, pp. 1817-1823, 1999.

The "patholog" of the gene expression profile in evaluating the ecotoxins effects

A.D. Iacobaş[1,2], M. Urban[2], S. Iacobaş[3] & D.C. Spray[2]

[1] Dept. Biophys. & Biomath., Ovidius University, Constanţa, Romania
[2] Dept. Neurosci. A.Einstein College of Medicine, New York, NY
[3] Dept. Pharmacology, Ovidius University, Constanţa, Romania

Abstract

The "spotting" microarray technique, consisting in large sets of DNA sequences spotted on poly-L-lysine-coated glass microscope slides, was developed to analyze genome-wide patterns. It is a tool to quantitatively monitor gene expression profiles and to analyze the alterations produced by genetic diseases, or induced by treatments, abnormal nutrition, and toxins. The pre-Hilbert space of standard gene expression (SSGE), based on the normal variability in gene profiles of each population from the ecosystem, is introduced. In this space, every point represents a possible gene profile and every continuous curve a possible genetic evolution. The gene clusters are identified by the correlation coefficients and the prominent gene of every cluster selected from the criteria of highest connective power and smallest variability. We define the gene "patholog" as the Euclidian distance separating the representative point of the population from the unit, 0-centered hypersphere. It is the most general global quantification of the alteration in digital genes expression, suitable to evaluate the effects of various toxins at the level of any particular population. A publicly available database will be open where every experimentalist could introduce his/her results and process them according to our procedure.

1 Introduction

Practically, any severe disease, nutrition deficiency, drug abuse, and long term action of toxins modify the genes' expression level, i.e. the amount of synthesized mRNA. This alteration can be now determined by the revolutionary technique of spotting microarray [1,2,3,4]. It seems to be a valuable indicator for medical [5,6,7], pharmacological [8,9,10], and (why not?) ecological investigation.

The acquisition and mathematical processing of the microarray images are difficult tasks since large amount of non-independent data have to be handled [11], [12], [13]. Packages of software for microarray results acquisition and analysis are now available (e.g. ScanAlyze, written by Eisen [14] and GenePix Pro 3.0 of Axon Instruments, Inc.), everyone trying to minimize the errors and increase the obtainable amount of information. [15], [16]

Several aspects of the microarray method could and should be further improved, part of them related to the calibration procedure [17] and standardization , others to the data interpretation [18]. We proposed recently some additional experiments and extra mathematical calculi to the common processing protocol [19], as done by ScanAlyze package, to: get the expected values and controllable fluctuations of the gene expressions [20], find the gene clusters [21] and the prominent gene of each cluster, build the pre-Hilbert space of standard gene expressions (SSGE) [22], [23], and compute the "patholog" of the gene expression profile. [24], [25]. The contributions of our group make one able to classify the genetic diseases and their stages according to the value of the patholog [26], [27]. The first application has been to analyze the modification of the genetic profile in the (mouse) N2A cells, stable transfected with the connexins Cx36 (which seems to accelerate cell differentiation) and Cx50 (which seems to retard the process) [28], [25]. This example will illustrate how to use our method in ecotoxicological gene based studies.

We consider the microarray technique as improved by our group and the philosophy of the Theory of the (here gene profile) patholog [29], [30] suitable to be extended at the study of toxin effects at the level of plant, animal, or human populations. It is a deep interest to know and prevent the alteration in genes expression level before a mutation is produced, and to have strong criteria to evaluate and optimize the action/treatment used in ecological reconstruction of the territory [31], [32].

This contribution presents for the first time the bases of our method for ecotoxicology. We have not yet enough experimental values to test it since large-scale studies have to be performed. Instead, we have the possibility to prepare and process the chips of the others who might be interested to further exploit their data in our way. Therefore, we are ready to open a public database and to co-operate with any interested group.

2 The space of standard gene expressions

It is a matter of evidence that: 1) the genes have different importance for the development of any tissue/organ/system; 2) the expression level of one gene depends generally on the expression level of some other genes [10], and 3) the genes are organized in (partially unknown) clusters [7], [15]. Therefore, we have built the poly-dimensional pseudo-pre-Hilbert space of standard gene expressions (SSGE) [13], based of the normal tissue variability [20], where every point represents a possible genetic situation and every continuous curve a possible evolution. Then, the gene clusters are identified based on gene correlation coefficients and the most prominent gene is selected from every cluster. This is a new and original way to

analyze the microarray data, introducing rigorous criteria in evaluating the genetic alterations and the efficiency of various treatments. Similar approaches has been used when evaluating and optimizing by computer simulation the medical care system in many other works [33], [34], [35].

Let x_j be the expression ratio (versus the normal/reference tissue in the green channel [17]) of the gene number j within the investigated genome map (containing U genes), $<x_j>$ the ratio when hybridizing normal tissue versus normal tissue, and σ_j the standard deviation (considered as controllable fluctuation in the normal tissue [36], [37]. $<x_j>$ is not 1 because of differences in cDNA labeling and spot capture, as discussed in [16], [19]. In the experiments we performed until now, [19], [24], [25], $<x_j>$ generally exceeded 1 and σ_j was below 1, for all investigated genes if the reference is in the green channel (see for instance Figure 1).

Continuing the research of Dr. Spray's group from AECOM, concerning the gap junction and the action of connexins [28], we analyzed the genetic profile of randomly selected 368 mouse genes from N2A normal cells (see Figure 1) and from those stable transfected with Cx36, processing the results by our original method. The method can be extended for any number of individual genes and any number of investigated organisms to fill a large database. We studied also 1158 human genes.

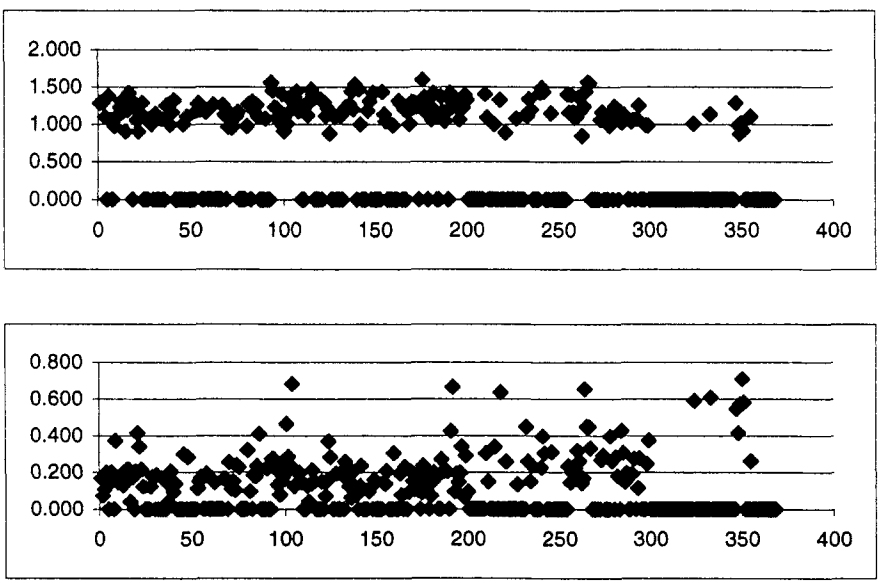

Figure 1: The average ratios (up) and the Student 5% maximum error of estimate (down) for 368 randomly selected mouse genes in normal cells. Differently labeled parts of the same source of cDNA, obtained from NA2 cells have been hybridized on 5 test chips from the same spotting series. 0-ratio is for the not expressed genes. The *norm* between the two channels ("red" and "green") had been set to 1.

Transforming the gene expression ratios into standard scores [36], one gets the possibility to compare them, whatever be their relative importance. This transformation is justified by the natural assumption that:

more important is one gene for the development of the organism at a particular moment in the cell life, less extended is its interval of controllable fluctuation.

By evidence, when normally expressed, the standard score of the gene expression is within the interval *(-1, +1)*.

The Cartesian product of the ranges of the standard scores can be formally organized as a continuous poly-dimensional pseudo-pre-Hilbert space, hereafter denoted by *Z* (see Figure 3), having as many dimensions as many individual genes are considered. It is pre-Hilbert since the norm derives from a scalar product and a pseudo-space since its unit vectors basis is not orthogonal [37]. In this mathematical space, every point, described by the poly-dimensional pseudo-vector $\vec{z} : z_1, \cdots, z_U$,

$$\vec{z} \equiv \begin{pmatrix} \dfrac{x_1 - < x_1 >}{\sigma_1} \\ \cdots \\ \dfrac{x_U - < x_U >}{\sigma_U} \end{pmatrix} \in \begin{pmatrix} \left(-\dfrac{< x_1 >}{\sigma_1}, \dfrac{\sum_{k=1}^{U} < x_k > - < x_1 >}{\sigma_1} \right) \\ \cdots \\ \left(-\dfrac{< x_U >}{\sigma_U}, \dfrac{\sum_{k=1}^{U} < x_k > - < x_U >}{\sigma_U} \right) \end{pmatrix} \equiv Z \tag{1}$$

represents a possible genetic expression profile and every continuous smooth curve a possible evolution. The ranges are not symmetrical because the minimum expression ratio of one gene is *0* and the maximum is the sum of all gene expressions, as in the pure hypothetical case that one and only one gene is transcribed at a moment and it is a limited amount of mRNA that can be present in the cell at one time.

The points at the borders of each range are critical (between possible and impossible expression). The situation is more critical if more scores are at the limit. Therefore, only the subspace inscribed in this Cartesian product, hereafter denoted by *Z*, which excludes the corners, is acceptable for our purpose.

Apparently, the hypercube defined by:

$$H = \left\{ \vec{z} \mid \vec{z} \in Z, \left| z_j \right| \leq 1 \quad , \quad \forall j = \overline{1, U} \right\} \tag{2}$$

is the set of all normally (here genetically expressed) situations. The problem is again at the borders of this hypercube, which represent critical states between normal and abnormal expression. We proved elsewhere by irreversible thermodynamics [22] that the inner of the 0-centered unit hypersphere *S* (the one inscribed in the above defined hypercube) contains all possible normal and stable states:

$$S = \left\{ \vec{z} \mid \vec{z} \in Z, \left\| \vec{z} \right\| \leq 1 \right\} \tag{3}$$

where $\left\| \vec{z} \right\|$ is the norm of the vector inside.

3 The gene clusters

There is enough evidence about the existence of non-linear (branched) gene networks [7]. Knowing these clusters will increase our understanding of genetic message and we shall get strong criteria to classify the diseases involving alterations in gene expression. This is a very laborious task, requiring many experiments in various conditions and the development of powerful mathematical and informatics tools. To find the gene clusters means to group the genes according the range of the correlation coefficients, hereafter denoted by the ρ's, between their expression ratios in various extracts. Then, one can study the properties of the gene network thus formed, building the connection graph and computing its power. For ecotoxicological purposes, it is important to know how the genes are correlated since the ecologist could act indirectly on a particular one through the effect produced on another, more accessible, gene. This knowledge is also important to predict the full effect of toxins at the population level.

Suppose two distinct genes denoted by their spot numbers in the microarray, i and j, and a set of r extracts from the same type of tissue. All extracts are analyzed versus the same reference and the results of the two spots compared. The correlation coefficient ρ_{ij} [36], [37] between the two genes is by definition:

$$\rho_{ij} = \frac{\mathrm{cov}(x_i, x_j)}{\sqrt{V(x_i)V(x_j)}} = \frac{\frac{1}{r}\sum_{k=1}^{r}\left[\left(x_i^{(k)} - < x_i >\right)\times\left(x_j^{(k)} - < x_j >\right)\right]}{\sqrt{\left(\frac{1}{r}\sum_{k=1}^{r}\left(x_i^{(k)} - < x_i >\right)^2\right)\times\left(\frac{1}{r}\sum_{k=1}^{r}\left(x_j^{(k)} - < x_j >\right)^2\right)}} \tag{4}$$

where $x_i^{(k)}$, $x_j^{(k)}$ are the ratios for the spots i and j in the chip k ($k = 1, 2, \ldots r$), and $<x_i>$ and $<x_j>$ are the two simple arithmetic averages of the spots i and j.

If two genes are positively correlated, then the extract with a higher expression in one is expected to manifest a higher expression in the other too. If negatively correlated, then a higher expression in one will diminish the expression in the other.

The representative gene of the cluster will be selected by the following criteria:

- *has the highest sum of square correlation coefficients within the cluster*
- *has the smallest confidence interval.*

The first criterion shows its influence in regulating the activity of the other genes from the cluster, while the second shows its importance, since we mentioned above, it is quite natural to consider that *more important is a gene, less liberty to fluctuate (even controllable) it has.* Therefore, the gene i connective power, $GCP(i)$, in the cluster (see Figure 2), C, defined by the composed gene numbers within the microarray, is:

$$GCP(i) = \sum_{j \in C} \rho_{i,j}^2 \tag{5}$$

while the percent gene i prominence, $GP(i)$, is:

$$GP(i) = \frac{1}{o_i} \times \frac{\sum_{j \in C} \rho_{i,j}^2}{\sum_{\substack{i,k \in C \\ k>i}} \rho_{i,k}^2} \times 100\% \tag{6}$$

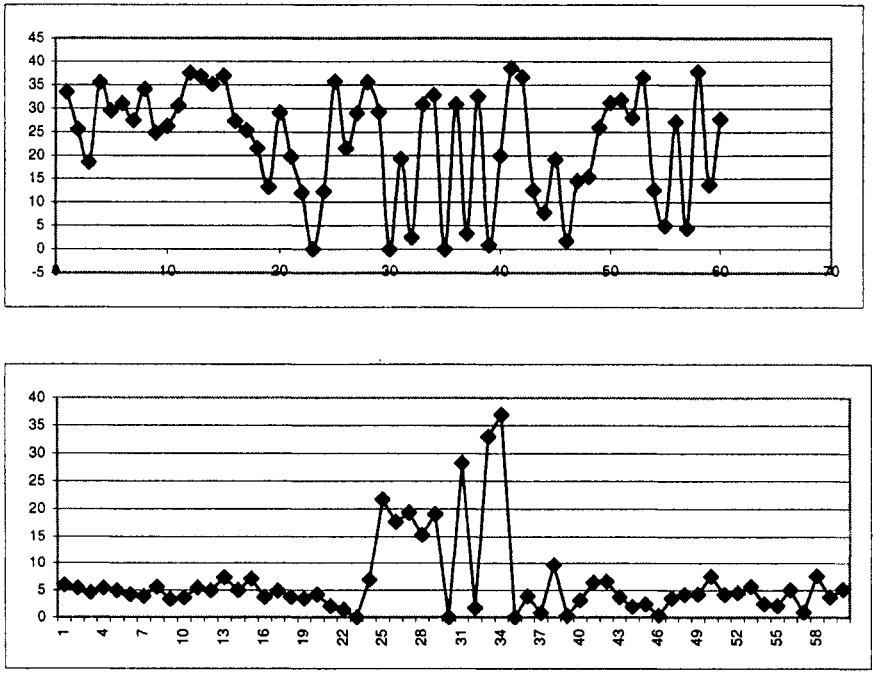

Figure 2: The gene connective powers (up) and the percent gene prominences (down) considering the clusters formed with 60 out of 368 randomly selected mouse genes, stable transfected with Cx36. The separating genes are on spots 23, 30, and 35, while the representative genes of the well-defined clusters are on spots 25 and 34.

4 The "patholog" of the gene expression alteration

Since the genes have different importance for the further evolution of the organism, and to include in the reports the complete list of gene expression modifications as a result of a toxin makes very difficult the evaluation, it is necessary to define a global indicator to characterize the entire set of the analyzed genes. We introduced the value of the "patholog of the gene expression" for various genetic transcription alterations as the Euclidian distance separating the representative point in the SSGE from the 0-centered unit hypershere, which is the subspace of normal and stable gene expression profile. (see Figure 3).

$$P(\vec{z} \in Z) = \begin{cases} d(\vec{z}, S) & , \quad if \ \vec{z} \notin S \\ 0 & , \quad if \ \vec{z} \in S \end{cases} \tag{7}$$

where S is the subspace defined in eqn. (3) and d defines the Euclidian distance between one point and a subspace in a poly-dimensional space.

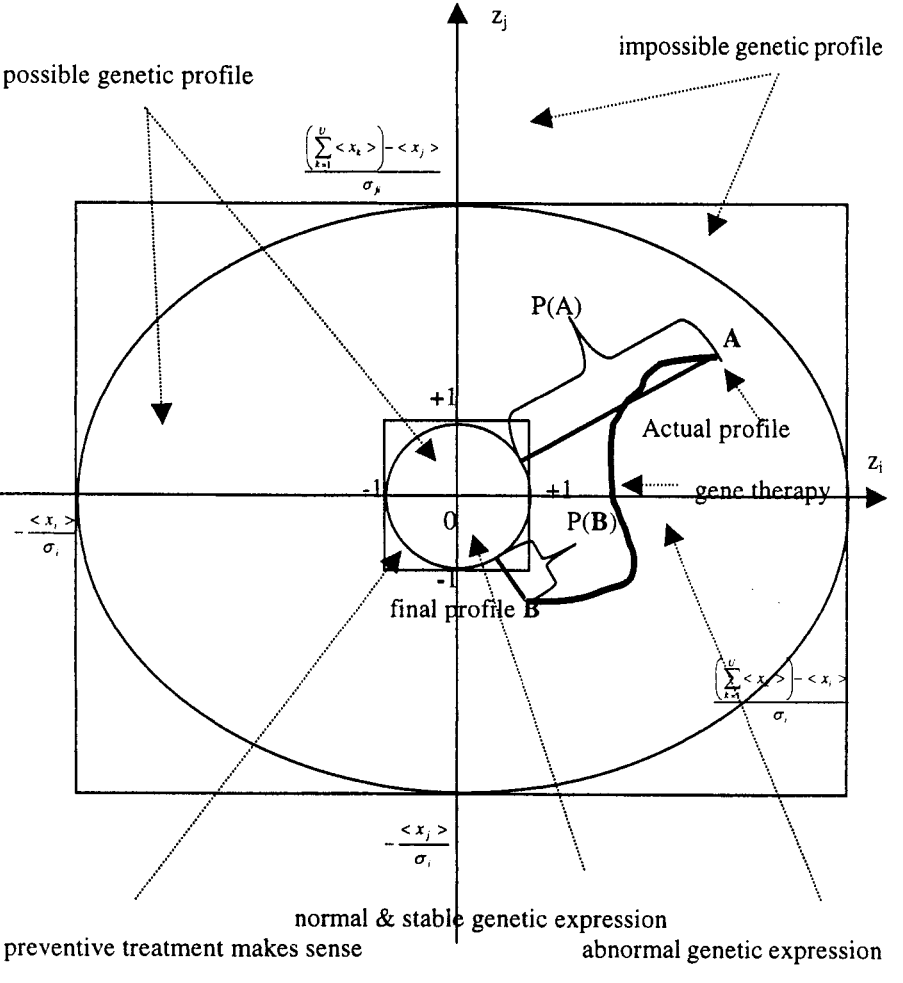

Figure 3: Two-dimension section through the poly-dimensional pre-Hilbert space of standard gene expression. The arrows show the important subspaces. **A** is the actual state and **B** the state after the treatment. The distances separating them from the 0-centered unit hypresphere are the pathologs of the genetic expression, P, for these states.

Note: Preventive action/treatments make sense only if the representative point is within the hypercube but not within the hypersphere [21]. When outside the hypercube, the system needs "therapy" because an ecological "pathology" has been already installed. When inside the 0-centered unit hypersphere, the system is under normal and stable ecological conditions. Thus, the manager of the ecosystem can evaluate the utility of a particular action to restore the normal ecological conditions.

5 Genetic profile alteration

Figure 4 presents the average standard scores of the gene expression levels in stable transfected N2A cells with Cx36. The background has been subtracted using the mean intensity of the pixels outside the spots.

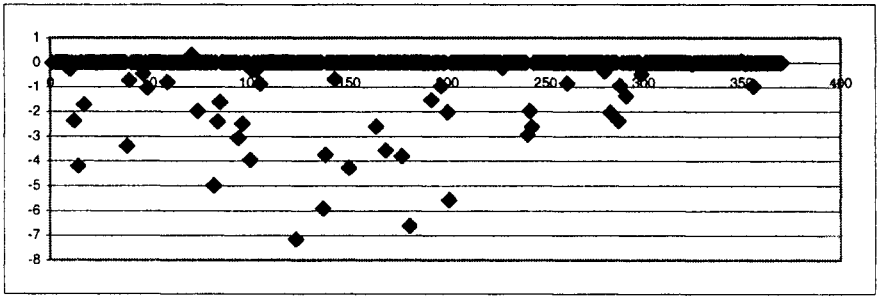

Figure 4: The standard scores of gene expressions for 368 randomly selected mouse genes in transfected N2A cells with Cx36. The patholog of this genetic profile is P = 19,463.

6 Perspectives instead of conclusions

By a similar procedure as presented here, one can compute the standard scores of gene expression levels and the patholog of any genetic profile, for any altering agent (as toxin) and population size. Then, one can determine a scale of patholog values, every interval imposing a particular type and cost level of action. Anyhow, whatever be the genes whose expression has been altered, for the same value of the patholog, the situation presents the same risk for the behavior of the ecosystem. Therefore, one can introduce some standards and criteria in evaluating and optimizing the large-scale actions.

References

[1] Brown P.O., Botstein D. Exploring the new world of the genome with DNA microarrays. *Nature genetics supplement.* 21: 33-37, 1999.

[2] Cheung V.G. Morley M., Aguilar F., Massimi A., Kucherlapati R., Childs G. Making and reading microarrays. *Nature Genetics 21 (1 Suppl):* 15-19, 1999.

[3] Eisen M.B., Brown P.O. DNA arrays of gene expression. *Methods in Enzymology*, 303: 179-205, 1999.

[4] Evertsz E., Starink P., Gupta R., Watson D. Technology and applications of gene expression microarrays. In *Microarray Biochip Technology*, ED. M.Schena, Eaton Publishing. Natick, MA. pp. 149-166, 2000.

[5] Audic S., Claverie J.M. The significance of digital gene expression profiles. *Genome Res.* 7(10), 986-995, 1998.

[6] Iacobaş A.D., Iacobaş S. Noi concepte şi indicatori în terapia genetică. *Analele Universităţii Ovidius. Seria Ştiinţe medicale.* **7**, 41-47, 2001.

[7] Tavazoie S., Hughes J.D., Campbell M.J., Cho R.J., Church G.M. Systematic determination of genetic network architecture. *Nature genetics.* 22: 281-285, 1999.

[8] Iacobaş A.D., Iacobaş S. Noi indicatori farmacologici în terapia genetică. *Farmacia* 1, 2001 (to appear).

[9] Iacobaş A.D., Urban M., Iacobaş S., Spray D.C. Premizele unei clasificări farmacologice pe baza pattern-urilor profilelor de exprimare a genelor individuale. *Farmacia.* 1. 2001 (to appear)

[10] Scherf U., Ross D.T., Waltham M., Smith L.H., Lee J.K., Tanabe L., Kohn K.W., Reinhold W.C., Myers T.G., Andrews D.T., Scudiero D.A., Eisen M.B., Sausville E.A., Pommier Y., Botstein D., Brown P.O. A gene expression database for the molecular pharmacology of cancer. *Nature genetics*, 24: 236-244, 2000.

[11] Bittner M., Meltzer P., Trent J. Data analysis and integration: of steps and arrows. *Nature genetcs.* 22: 213-215, 1999.

[12] Gaasterland T., Bekiranov S. Making the most of microarray data. *Nature Genetics*, 24: 204-206, 2000.

[13] Iacobaş A.D., Urban M., Iacobaş S., Spray D.C. Bazele matematice şi conceptuale ale analizei pattern-urilor profilelor de exprimare ale genelor individuale. *Analele Universităţii Ovidius. Seria Ştiinţe medicale.* 7, 48-52, 2001.

[14] Bassett Jr. D.E., Eisen M.B., Boguski M.S.. Gene expression informatics – it's all in your mine. *Nature genetics. Supplement.* 21: 51-55.

[15] Smolen P., Baxter D.A., Byrne J.H. Modeling transcriptional control in gene networks – methods, recent results, and future directions. *Bull. Math. Biol.* 62, 247-292, 2000.

[16] Spray D.C., Iacobaş A.D., Urban M. Theoretical and practical optimization of gene microarray technique. *Medinf'2000.* (www.umfiasi.ro/medinf)

[17] Basarsky T., Verdnik D., Zhai J.Y., Wellis D. Overview of a microarray scanner: design essentials for an integrated acquisition and analysis platform. *Microarray Biochip Technology*, ed. M.Schena, Eaton Publishing. Natick, MA. pp.265-284, 1999.

[18] Carulli J.P., Artinger M., Swain P.M., Root C.D., Chee L., Tulig C., Guerin J., Osborne M., Stein G., Lian J., Lomedico P.T.. High throughput analysis of differential gene expression. *J. Cell. Biochem. Suppl.* 30/31: 286-296, 1998.

[19] Iacobaş A.D., Urban M., Spray D.C. New protocol in spotting microarray technique. *Rom. J. Physiology* 1-2, 2000. (to appear in 2001)

[20] Ross D.T., Eisen M.B., Perou C.M., Rees C., Spellman P., Iyer V., Jeffrey S.S., Rijn Mvd, Waltham M., Pergamenschikov A., Lee J.C.F., Lashkari D., Shalon D., Myers T.G., Weinstein J.N., Botstein D., Brown P.O. Systematic variation in gene expression patterns in human cancer cell lines. *Nature genetics.* 24: 227-235, 2000.

[21] Tortonen P., Kolehmainen M., Wong G., Carsten E. Analysis of gene expression data using self-organizing maps. *FEBS Letters,* 451: 142-146, 1999.

[22] Iacobaş S., Iacobaş A.D. Functions with medical meaning defined on a pre-Hilbert space of normal and pathological states of the human organism. *Actual Problems in Biophysics.* 7, 205-208, 1983.

[23] Iacobaş S., Iacobaş A.D. A mathematical transcription of a current problem in physiology. *PHYSIOLOGIE,* 24(4), 267-270, 1987.

[24] Iacobaş A.D., Iacobaş S, Spray D.C. The "patholog" of the genes expression profile, a new tool in defining, evaluating, and classifying the genetic diseases. *Rom. J. Physiology* 1-2, 2000. (to appear in 2001)

[25] Iacobaş A.D., Urban M., Iacobaş S., Spray D.C. The "patholog" of the genes expression profile in stable transfected mouse N2A cells with Cx36 and Cx50. *Ars Medica Tomitana.* VII(25) (in press)

[26] Iacobaş A.D. Cancer classification by analyzing the patterns in the pre-Hilbert space of gene expression. *MEDINF'2000* (www.umfiasi.ro/medinf). 2001.

[27] Iacobaş A.D., Iacobaş S. Mathematical set up of an actual medical problem. *Analele Universitatii "Ovidius" Constanţa. Seria Ştiinţe medicale.* II, 152-160, 1991.

[28] Rozental R, Srinivas M., Gokhan S., Urban M., Dermitzel R., Kessler J.A., Spray D.C., Meltzer M.F. Temporal expression of neuronal connexins during hippocampal ontogeny. *Brain Res. Rev.* 32: 57-71, 2000.

[29] Iacobaş A.D., Iacobaş S. Evaluation and validation of the health care system by the Theory of pathologic. (Chapter 47) *Advances in Medical Physics, Biophysics and Biomaterials.* Ed. E. Kukurova, Male Centrum, Bratislava, pp. 175-179, 1997.

[30] Iacobaş A.D., Iacobaş S. Drug efficiency estimate with the Theory of Pathologic. *MEDINF'98.* 80-85, 1998. http://medinfo.umft.ro/rsmi.

[31] Iacobaş A.D. New ideas in thermodynamical ecology. *Ecology of industrial regions.* 2(1-2), 96-100, 1996.

[32] Iacobaş Sanda, A.D.Iacobaş. Principles of ecological reconstruction. *Ecology of industrial regions.* 2(1-2), 101-106, 1996.

[33] Iacobaş A.D., Iacobaş S. Algoritm pentru optimizarea pe computer a tratamentelor medicale. (Chapter 14) *Cibernetica aplicată.* ed. M.Manescu. pp. 117-122, 1985.

[34] Iacobaş A.D., Iacobaş S. BASIC program in monitoring and analysing of a living system. *PHYSIOLOGIE.* 24(4), 267-270, 1987.

[35] Iacobaş S., Iacobaş A.D. Le pathologique – synthèse des informations sur l'organisme humain. (Chapter 12) *Aspects energetiques et informationels dans les systemes vivants*, ed. M. Godeanu. ICEBIOL, Bucharest. pp. 127-134, 1986.

[36] Iacobaş A.D. *Medical Biostatistics.* 3rd English edition *Bucura Mond,* Bucharest, pp. 1-33, 85-89, 1997.

[37] Iacobaş A.D. *Ideas and methods in the physics of living.* 4th English edition, Tilia Press International, Constanţa, pp. 273-300, 2000.